CHARLES UNIVERSITY

Faculty of Science

Department of Biochemistry

Study program: Biochemistry

Mgr. Jan Fiala

Development of pulse labelling technology for studying the dynamics of protein complexes.

Vývoj technik pulsního značení pro studium dynamiky proteinových komplexů.

DOCTORAL THESIS

Supervisor: RNDr. Petr Novák, Ph.D.

Advisor: RNDr. Petr Pompach, Ph.D.

Prague 2021

DECLARATION

I declare that I have worked on this thesis under the guidance of my supervisor and that all sources of the previous knowledge are properly cited. No part of this work was used or will ever be used to obtain any other academic degree than Ph.D. from Charles University.

Prague ……

……

Mgr. Jan Fiala

DECLARATION OF AUTHORSHIP

I declare that Mgr. Jan Fiala contributed significantly to the experiments and to all scientific publications contained in this Ph.D. thesis. He performed most of the experiments, substantially contributed to their planning, and took a significant part in the primary data interpretation and their preparation for publication.

Prague ……

……
RNDr. Petr Novák, Ph.D.

ACKNOWLEDGEMENTS

Herein, I would like to express my immense gratitude to all those who accompanied me throughout my Ph.D. studies.

First of all, I would like to thank my supervisor RNDr. Petr Novák Ph.D. for his patience in leading this work and all the time and willingness he devoted to me. Also, I want to stress out I was delighted by his motivating approach to solving experiments, by providing me the opportunity to participate in several exciting projects, by his valuable advice and especially by his boundless patience.

Many thanks also go to all the other members of the Laboratory of Structural Biology and Cell Signaling, who have become more of a second family during the time. Especially to Petr Man, Petr Pompach, Petr Halada, Zdeněk Kukačka, Aleš Hnízda for their invaluable practical advice and almost endless scientific conversations, and to Lukáš Fojtík for sharing an emotion from all experiments. Last but not least, enormous thanks belongs to Tereza Kadavá and the Appendix of the "jontová mobilička skupina" for the countless hours of late-night measurement.

I would also like to thank our collaborators, especially Daniele Fabris from the University of Connecticut, for the unique opportunity of a six-month internship in his laboratory and for providing me insight into the utterly unknown field of native mass spectrometry. Further thanks go to Sergei Strelkov from Katholieke Universiteit Leuven, who was in charge of the CELSA project, partially revealing the beauty of intermediate filaments assembly.

However, I would most like to thank my beloved wife, Jana Fialová, who constantly supports me and gives me the courage to walk towards my dreams. Last but not least, to all my family members for their endless support during my never-ending studies.

ABSTRACT (IN ENGLISH)

Structural mass spectrometry (MS) is an evolving field of structural biology introducing novel techniques for the characterization of biomolecules. Although MS-based techniques only can provide "low-resolution" information compared to standard high-resolution techniques representing by X-ray crystallography, cryo-electron microscopy or nuclear magnetic resonance, its uniqueness lies in the ability to easily obtain structural information about various biomolecules in their native or native-like environment.

By employing various approaches, from protein covalent labelling through chemical cross-linking to ion mobility, structural MS provides insight into the structure and dynamics of proteins and their complexes over a broad timescale.

This thesis is dedicated to the development of novel structural MS approaches based on pulse covalent labelling and chemical cross-linking. Employing the developed quench-flow microfluidics apparatus, we performed footprinting experiments on proteins and protein complexes in timescale from a few microseconds to single seconds.

Specifically, fast photochemical oxidation of proteins (FPOP) and novel fast fluoro alkylation of proteins (FFAP) techniques were utilized to track structural changes of myoglobin upon release of the prosthetic heme group. Additionally, these footprinting techniques were successfully used to localize the interaction interphase of the biologically important human haptoglobin-haemoglobin complex, which atomic structure was not still satisfactorily described. Further, chemical cross-linking was beneficially used to describe the longitudinal interaction of "mini-lamins" tetrameric assembly.

KEYWORDS

Structural mass spectrometry, proteins and protein complexes, protein dynamics, higher-order structure (HOS), pulse labelling, covalent labelling, chemical cross-linking, radical footprinting, fast photochemical oxidation of proteins (FPOP), Togni reagents, fast fluoroalkylation of proteins (FFAP).

ABSTRAKT (IN CZECH)

Strukturní hmotnostní spektrometrie (MS) je progresivní metoda současné strukturní biologie pro charakterizaci biomolekul. Ačkoli techniky založené na MS poskytují pouze informace "s nízkým rozlišením" ve srovnání s klasickými strukturálními metodami, jako je rentgenová krystalografie, elektronová mikroskopie nebo nukleární magnetická rezonance, její jedinečnost spočívá ve schopnosti snadno získat strukturní informace o různých biomolekulách v jejich nativním prostředí.

Využitím různých přístupů, od kovalentního značení proteinů, přes chemické síťování až po iontovou mobilitu, umožňuje strukturní MS pohled na strukturu a dynamiku proteinů a jejich komplexů v čase.

Tato práce je věnována vývoji nových přístupů strukturní MS, které jsou založeny na pulzním kovalentním značením a chemickém síťování. S využitím vyvinutého mikrofluidního reaktoru byl proveden "footprinting" povrchu proteinů a proteinových komplexů v časovém rozsahu od několika mikrosekund až po jednotky sekund.

V této práci byly konkrétně použity rychlá fotochemická oxidace proteinů (FPOP) a nově vyvinutá technika rychlé fluoroalkylace proteinů (FFAP) pro sledování strukturních změn myoglobinu způsobených uvolněním prostetické skupiny hemu. Dále byly tyto metody úspěšně aplikovány k lokalizaci interakčního rozhraní biologicky důležitého komplexu lidského haptoglobinu a hemoglobinu, jehož atomová struktura nebyla dosud uspokojivě popsána. Dále bylo použito chemické sítění k popisu podélné interakce tetramerního komplexu "mini-laminů".

KLÍČOVÁ SLOVA

Strukturní hmotnostní spektrometrie, proteiny a proteinové komplexy, strukturní dynamika, struktura vyššího řádu (HOS), pulzní značení, kovalentní značení, chemické zesítění, radikálové značení, rychlá fotochemická oxidace proteinů (FPOP), Togniho činidla, rychlá fluoralkylace proteinů (FFAP).

TABLE OF CONTENTS

ABBREVIATIONS

3D	three-dimensional
BS2G	bis(sulfosuccinimidyl) glutarate
BSA	bovine serum albumin
CC	coiled-coil structure
CCS	collision cross-sections
CD	circular dichroism
cIMS	cyclic ion mobility spectrometry
CIU	collision-induced unfolding
CL-MS	cross-linking mass spectrometry
Cryo-EM	cryogenic electron microscopy
CXMS	cross-linking mass spectrometry
DIMS	differential ion mobility spectrometry
DMA	differential mobility analyzers
DMF	dimethylformamid
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
DSBU	disuccinimidyl dibutyric urea
DSG	disuccinimidyl glutarate
DSPU	disuccinimidyl dipropionic urea
DTIMS	drift tube ion mobility spectrometry
ECD	electron capture dissociation
EDC	*N*-ethyl-3-*N'*,*N'*-dimethylaminopropylcarbodiimide
ESI	electrospray ionization
ETD	electron transfer dissociation
FAIMS	field-asymmetric waveform ion mobility spectrometry
FFAP	fast fluoroalkylation of proteins
FGF	fibroblast growth factor
FID	free-induction decay
FPOP	fast photochemical oxidation of proteins
FT	Fourier transformation
FT-ICR	fourier transform ion cyclotron resonance

Hb	hemoglobin
HDX	hydrogen deuterium exchange
HOS	higher-order structure
Hp	haptoglobin
IAA	iodoacetamide
IFs	intermediate filaments
IM	ion mobility
IMS	ion mobility spectrometry
IMS-MS	ion mobility spectrometry-mass spectrometry
IR	infrared radiation
IRMPD	infrared multiphoton dissociation
KrF	krypton fluoride
LC-MS	liquid chromatography mass spectrometry
LCP	left circularly polarized
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
multiCASI	multiple continuous accumulation of selected ions
Nd:YAG	neodymium-doped yttrium aluminum garnet
nESI	nano electrospray ionization
NHS	*N*-hydroxysuccinimide
NMR	nuclear magnetic resonance
PASEF	parallel accumulation−serial fragmentation
PBS	phosphate-buffered saline
PDB	protein data bank
QTOF	quadrupole time-of-flight
QXL-MS	quantitative cross-linking mass spectrometry
RCP	right circularly polarized
RF	radio frequency
ROS	reactive oxygen species
RPLC	reverse-phase liquid chromatography
SASA	solvent accessible surface area
SDA	succinimidyl 4,4'-azipentanoate
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SEC-DIS	size-exclusion chromatography coupled with dynamic light scattering
SID	surface-induced dissociation
SPB	succinimidyl-[4-(psoralen-8-yloxy)]-butyrate
Sulfo-NHS	*N*-hydroxysulfosuccinimide
TBHP	*t*-butylhydroperoxide
TCEP	tris(2-carboxyethyl)phosphine
TEM	transmission electron microscopy
TIMS	trapped ion mobility spectrometry
TWIMS	travelling wave ion mobility spectrometry
UV	ultraviolet radiation
UV-CD	ultraviolet circular dichroism
UVPD	ultraviolet photodissociation
vT-ESI	variable-temperature electrospray ion source
XL-MS	cross-linking mass spectrometry
XRC	X-ray crystallography

1. Introduction

1.1 Proteins and protein complexes

1.1.1 Structure and dynamics

The proteins are the essential and the most diverse biomolecules of life and underpin of biochemistry as a field. Such biomolecules were discovered in 1838 by Mulder[1] and subsequently named proteins (from the Greek πρώτα "prota") by Berzelius[2], meaning "primary", "in the lead", or "standing in front"[3]. Thenceforth, proteins have become pivotal molecules of research interest, revealing their irreplaceable role in biological processes such as cellular metabolism, cell growth, communication, transport, storage and many others.

However, proteins studies experienced the most significant expansion in the 1950s, when the fundamental phenomenon of linking amino acids into one long chain (primary structure) followed by hydrogen-bonding stabilization (secondary structure) was discovered[4,5]. Since then, the protein structure has become one of the major milestones of structural biology, which leads to the development of novel techniques, determining protein structures at all hierarchy levels known as primary, secondary, tertiary, quaternary structure (**Figure 1**)[3].

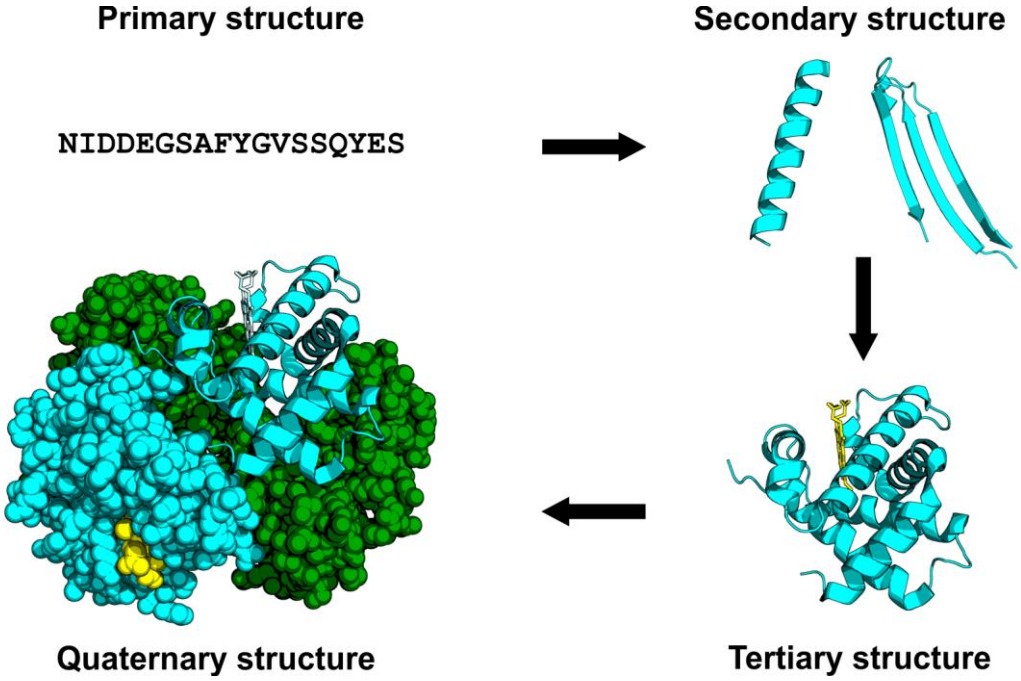


Figure 1: *Protein structure hierarchy illustrated on tetrameric haemoglobin $\alpha_2\delta_2$ (PDB: 1nqp[6])*

The first level of the structure hierarchy is the primary structure, which refers to the direction-dependent (N-terminus, C-terminus) sequence of linked amino acids serving as the basic building blocks.

In general, each amino acid is composed of a central alfa carbon (C_α) linked to an amino ($–NH_2$) and a carboxyl (–COOH) groups that link single amino acids to chains through the peptide amide bond (CO-NH). Additionally, C_α is a covalently binding R group, representing 20 different proteinogenic side chains, giving amino acids unique properties such as size, shape, charge and polarity. On top of that, C_α forms a chiral center (19 of 20amino acids), creating a D or L stereoisomers of amino acids. However, only L-stereoisomers are presented in natural protein constructs[7]. By linking amino acids, the polypeptide chain is formed. The order of linked amino acids results in the sequence (primary structure) having unique properties.

However, this simple linking phenomenon of the primary structure was not deciphered until 1952, when Sanger sequenced the first protein (insulin) by his novel approach[4]. In the same decade, the proposed secondary structure was experimentally approved by hydrogen-deuterium exchange (HDX) experiments[5,8].

The secondary structure is another level in protein structure hierarchy related to the arrangement of a linear peptide sequence in space. The most important interactions responsible for stabilizing secondary structure are backbone hydrogen bonds, resulting in the formation of recognizable α-helices and β-sheets (**Figure 1**).

The α-helices are peptide structures made up of right-handed coiled amino acids, with one whole turn made up of 3.6 residues. On the contrary, β-sheets are composed of two or more peptide strands situated next to each other in either a parallel or antiparallel arrangement.

The dependence of peptide sequence on the occurrence of secondary structures was deeply studied. One of the pioneering work was published by Ramachandran *et al.*[9], which correlates torsion angles φ (around N-C_α) and ψ (around C_α-C) of each amino acid with secondary structure likelihood. In principle, the peptide bond has a partial-double-bond character due to the presence of delocalized electrons within the resonance structure (**Figure 2a**). This fact makes peptide bond (CO-NH) reluctant to any rotation giving it a rigid "plane-like" character (**Figure 2b**). As consequences, only bonds with angles marked as φ and ψ has rotating freedom of 360 degrees. However, in proteins, full rotation is limited by neighbour side chains causing atoms clashes due to the steric hindrance. Thus, by plotting, φ over ψ results in the Ramachandran plot for every amino acid (**Figure 2c**), revealing

regions representing preferred, preferably allowed, exceptionally allowed, and prohibited angles combinations. These regions can also be interpreted as a preferred secondary structure.

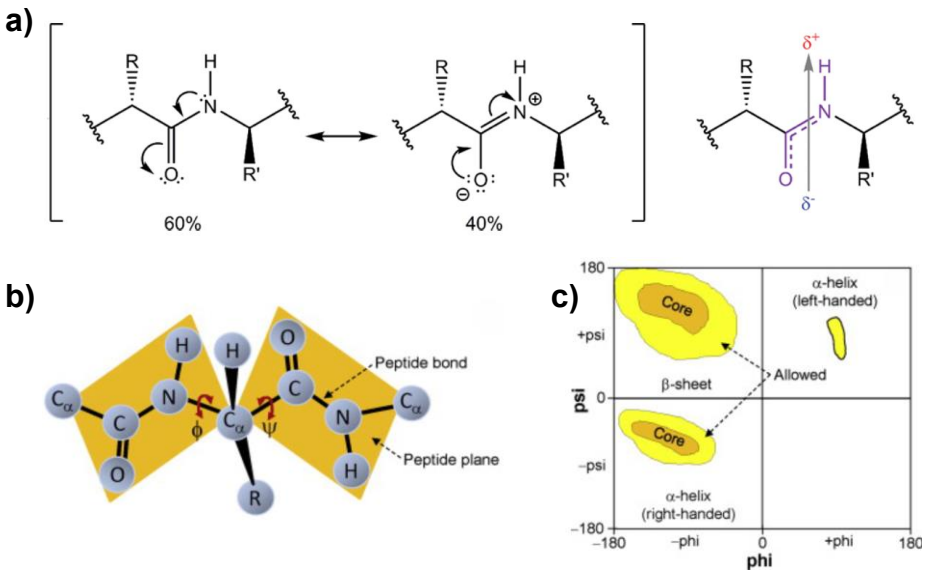


Figure 2: *a) Resonance structures of the peptide bond revealing its partial-double-bond character. b) Peptide with highlighted peptide plane with rotatable bonds N-C_α (φ) and C_α-C (ψ). c) Representative Ramachandran plot with regions of preferred (orange), allowed (yellow) and prohibited (white) angles combinations revealing secondary structure motifs. Adapted and edited from*[10,11].

Thanks to its simplicity, Ramachandran plots are ordinarily used to validate new structures during the deposition process in protein data bank (PDB)[12]. Additionally, they are used in bioinformatics for *in silico* predictions of protein structure and properties[11,13].

The tertiary structure is another protein level that refers to the 3D arrangement of previously described secondary structure motifs. The resulting domains (**Figure 3**) stabilized by hydrogen bonds, non-covalent interactions (Van der Waals interactions, ionic bonds), and disulfide bridges are already functional protein units capable of a particular function (such as binding cofactor, gene activation or catalysis). However, large proteins are usually composed of more than one domain, giving them a unique set of functions (e.g. activation domain and DNA-binding domain of transcription factors)[14].

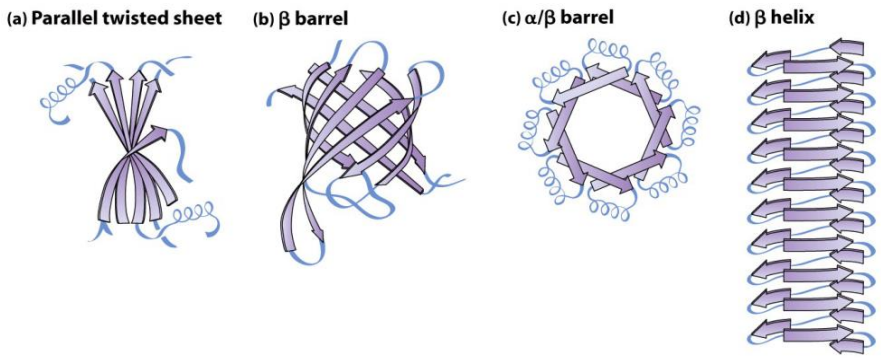


Figure 3: *Example of common domain folds. Adapted from*[15].

The additional level of protein arrangement is known as the quaternary structure, which is the organization of more than one polypeptide chain (subunit) into multiple subunits protein complex (oligomer). The oligomer is held together predominantly by weak hydrophobic non-covalent interactions (electrostatic interactions are also involved). Protein complex can be composed of identical or different subunits but always with defined stoichiometry.

For a description of the oligomer stoichiometry, Greek letters are used as descriptors of subunit type, and subscript numbers define a number of involved subunits. For illustration, the most common haemoglobin type is a tetramer composed of two α, and two β non-covalently bound subunits, which is denoted as haemoglobin $\alpha_2\beta_2$[15].

Since protein 3D structure and function are assumed to be in a close relationship, then the protein structure should be sufficient enough to thoroughly understand protein function. Nevertheless, structural information is providing just a static view of the system without its dynamics. However, native proteins undergo many conformational fluctuations associated with their stability, function, and regulation. Thus, understanding protein dynamics in a broad timescale is essential for understanding their essential functions[16].

One of the most critical dynamic events is protein folding which occurs simultaneously or immediately after translation of a new protein chain. Once the protein chain is synthesized as a random coil, its folding pathway is naturally driven into its native functional conformation. According to Anfinsen's dogma (also known as a thermodynamic hypothesis), native conformation is encoded in the protein's primary sequence and have one of the lowest free energy[17]. However, the protein's ability to spontaneously reach its native structure (lowest free energy state) in less than a few seconds is impossible from a thermodynamical point of view. Levinthal proposed this folding paradox in 1968, showing that finding native

conformation by a random search among all possible configurations would take an enormously long time[18,19]. Additionally, he suggested a paradox solution by the assumption that protein folding can be guided through local interactions, which determine further folding through the formation of nucleation points (**Figure 4a**)[18]. Therefore, protein folding can be illustrated as a "folding funnel" where protein occupies limited energetical levels corresponding to semi-stable structural folds (**Figure 4b**)[20].

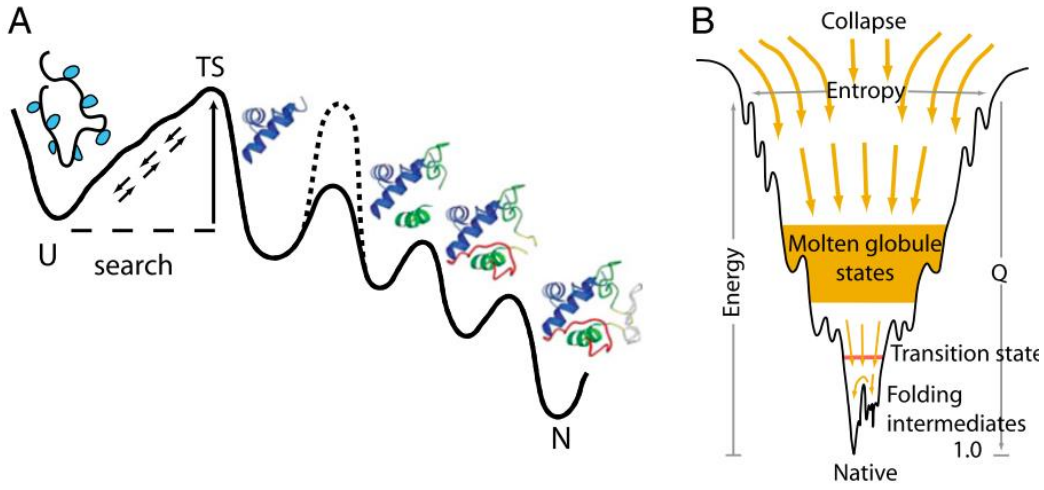


Figure 4: *a) Classic view representing protein folding pathway. b) Novel view of the folding scheme is a "folding funnel", where protein is subsequently undergoing its semi-stable folding intermediates through the funnel-like energetic landscape. Adapted from*[21].

1.1.2 Methods for the study of protein structures

1.1.2.1 Circular dichroism

Circular dichroism (CD) spectroscopy is a technique used for the determination of protein secondary structure. In general, the CD is based on distinct absorption of right and left-handed polarized light after passing through the protein sample[22].

In detail, light has time-dependent magnetic and electric field vectors oscillating perpendicularly to each other in the same phase (**Figure 5a**). By shifting magnetic or electric field phase by ¼, circularly polarized light is formed with its vector performing a circular motion (**Figure 5b**), yielding in the right circularly polarized (RCP) or left circularly polarized (LCP) light. Since proteins secondary structure is non-symmetric, it can absorb left or right-handed circularly polarized light differently (F**igure 5c**). Due to the absorption of one light component (decrease of amplitude), perfectly circular motion becomes elliptical (elliptically polarized see **Figure 5d**). The difference of these motions is expressed as (molar) ellipticity (θ) which is the unit seen in the CD spectrum measured over a range of wavelengths[23].

In the case of proteins, far-ultraviolet circular dichroism (UV-CD) spectra are measured due to the presence of characteristic absorption waveforms for α-helices and β-sheets (**Figure 5e**). Specifically, α-helixes have absorption maxima at positive 193 nm and negative 222 nm and 208 nm, yielding in characteristic negative double peak. On the other hand, β-sheets have positive 195 nm and negative 218 nm bands. Also, disordered proteins have recognizable maxima at positive 210 nm and negative 195 nm[22].

For the determination of the sample with unknown secondary structure, the measured spectrum is compared to reference spectra of fully α-helical or β-sheets proteins. By mathematical deconvolution, the presence of a relative fraction of the secondary structure can be determined. As was shown, the CD can provide only low-resolution information about protein structure. Hence, for detailed structural analysis, other techniques are involved and will be closely described[22].

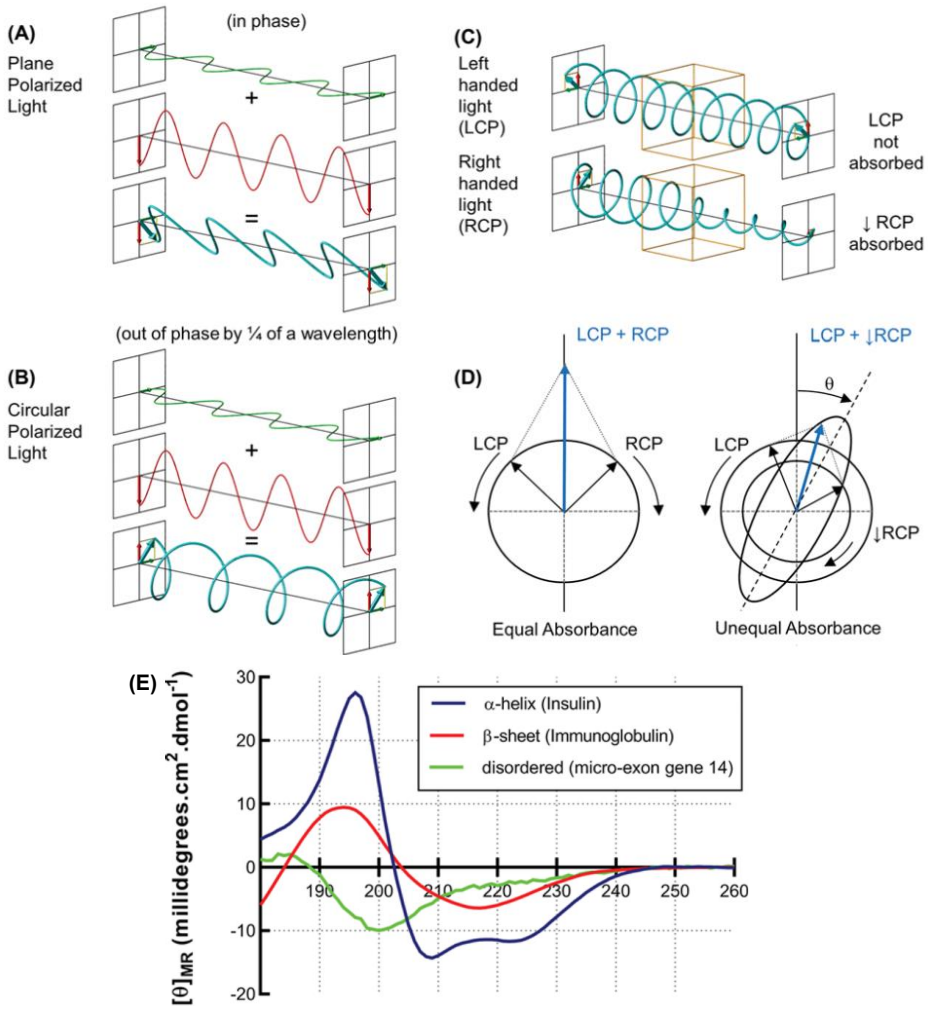


Figure 5: *Green and red waves oscillating perpendicularly to each other a) Plane-polarized light created by combining waves in the same phase oscillates back and forth at 45 degrees (represented in cyan). b) Circularly polarized light created by the combination of red and green waves off ¼ a wavelength phase. c) A chiral molecule (indicated as the red box) absorbing the only RCP light yielding in d) elliptically polarised light. e) UV-CD spectrum of various proteins: α-helical insulin (blue), β-sheeted immunoglobulin and disordered micro-exon gene 14 (green). Adapted from*[10].

1.1.2.2 X-ray crystallography

X-ray crystallography (XRC) is an analytical method studying the interaction of crystalline samples with X-ray irradiation. Unlike CD, XRC allows the determination of the absolute molecular structure by known positions of atoms, bond lengths and angles in the crystal lattice[24].

The first experiments revealing the feasibility of solving protein structures by X-ray crystallography dates back to 1958, when the first detailed protein structure of sperm whale myoglobin was uncovered by Kendrew e*t al.*[25]. Just after four years, Kendrew and Perutz were awarded by Nobel prize for their studies of the structures of globular proteins.

Subsequently, XRC has become the central technique for answering many biological questions resulting in more than ten additional Nobel prizes related to the utilization of XRC (most recent 2012; Structure and function of G-protein-coupled receptors)[26]. Until now, Protein Data Bank registered more than 178 000 macromolecular structure entries, whereas 157 000 were solved by X-ray crystallography. This fact is pinpointing XRC's fundamental role in structural biology[27].

In brief principle, during the pass of monochromatic X-rays through the substance, elastic bending (diffraction) of the X-rays occurs. The direction and intensity of the diffracting beams depend on the internal structure (electron density) of the sample. A periodic crystalline structure acts as a diffraction grating, resulting in the diffraction pattern (**Figure 6**) created by constructive/destructive light interferences determined by Bragg's Law. For obtaining complete protein structural information, the crystal must be scanned in many X-ray-to-crystal angles[24].

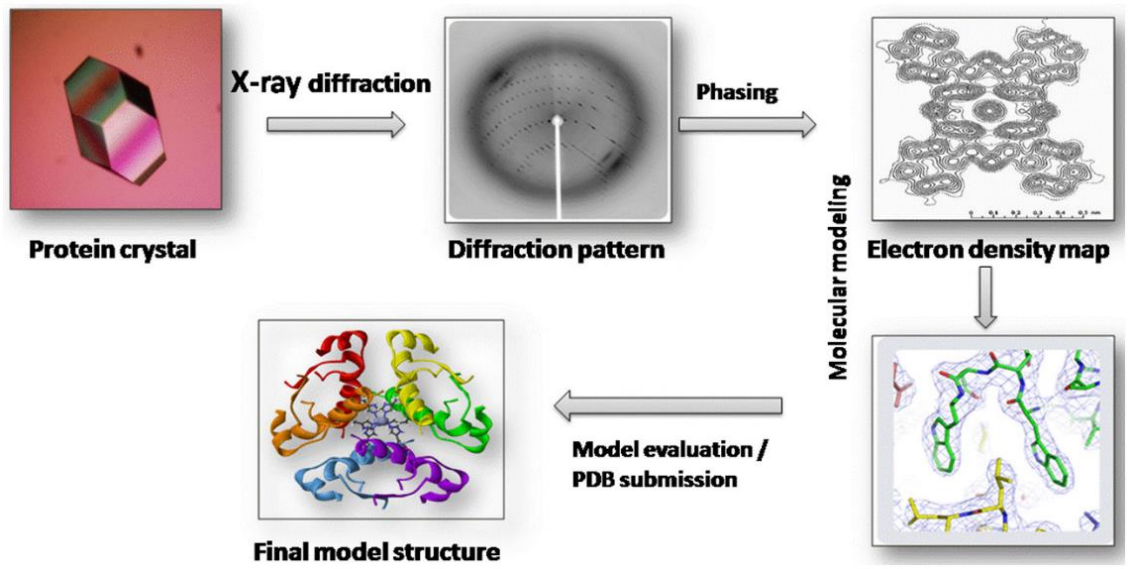


Figure 6: *Schematic workflow of X-ray crystallography experiment. Adapted from*[24].

The obtained diffraction patterns (reflections) are converted with Fourier transformation (FT), yielding in electron density map and vice versa. As amplitude, the square root of the spot intensity is used; however, the phase of the reflections is unknown and cannot be measured. Due to the missing phase information, FT cannot be directly applied, and electron density cannot be directly calculated. This phenomenon is known as a phase problem and can be solved with many approaches[10].

The most used approach is Molecular Replacement, where a similar structure is used to determine the unknown structure. In principle, the software calculates theoretical diffraction patterns of many orientations of known structure through FT followed by fitting theoretical reflections with experimentally measured data. Once the software finds an appropriate overlay, it adapts phases of the known model and creates the initial electron density map. Afterwards, the polypeptide chain with an unknown structure is fitted into the initial electron density map, and the FT calculates a theoretical diffraction map based on the position of fitted atoms. By comparing the newly calculated theoretical diffraction map and experimentally measured, the phases are improved, and a better fitting electron density map is created. This process named refinement is then looped until satisfactory fit is obtained. After refinement, the structure is deposited at PDB as x, y, z coordinates for each atom[10].

Even though X-ray crystallography proved its application for many biological relevant systems, its limitation lies in protein crystallization, which is not achievable within some proteins of interest[28]. Additionally, crystallized proteins provide only snapshot-like (frozen in time) structural information, and therefore XRC cannot be used to study protein dynamics. Therefore, other techniques such as nuclear magnetic resonance (NMR) or mass spectrometry have to be utilized[26].

1.1.2.3 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) spectroscopy is a method for studying the protein structure and its dynamic behavior, complexing abilities, and many other important physicochemical properties in solution. To obtain the 3D structure, the proteins must be isotopically enriched with magnetically active atoms (spin $\neq 0$) such as ^{1}H, ^{13}C, ^{15}N and ^{31}P. This can be achieved by expressing the respective protein in a suitable expression system (e.g. *Escherichia coli*) in an isotopically enriched growth medium[29].

Once the protein is isotopically enriched and placed in the strong magnetic field, the atom's nucleus magnetic moments (axis of rotation) arrange parallel to the external magnetic field B_0 (z-axis in **Figure 7**). Subsequently, a short radio frequency (RF) pulse generating variable magnetic field (B_1) perpendicular to the external field is applied, causing magnetic vectors to shift in (x-y plane). This shift is immediately followed by the return of excited magnetic moments in equilibrium by rotation at Larmor frequency (chemical shift in ppm). This signal is recorded ("echo") in the form of exponentially decaying current as time-dependent free-induction decay (FID)[30].

When proteins are measured, various atoms with many different Larmor frequencies are recorded at once, and Fourier transformation is needed to be applied to deconvolute NMR spectra to separate chemical shifts[30]. To determine protein 3D structure, 2D NMR experiments are usually conducted, providing additional information by recording magnetization transfer from one hydrogen to another in near 3D space. These experiments give clear evidence of the spatial arrangement of measured atoms used for molecular modelling and structure determination[31,32].

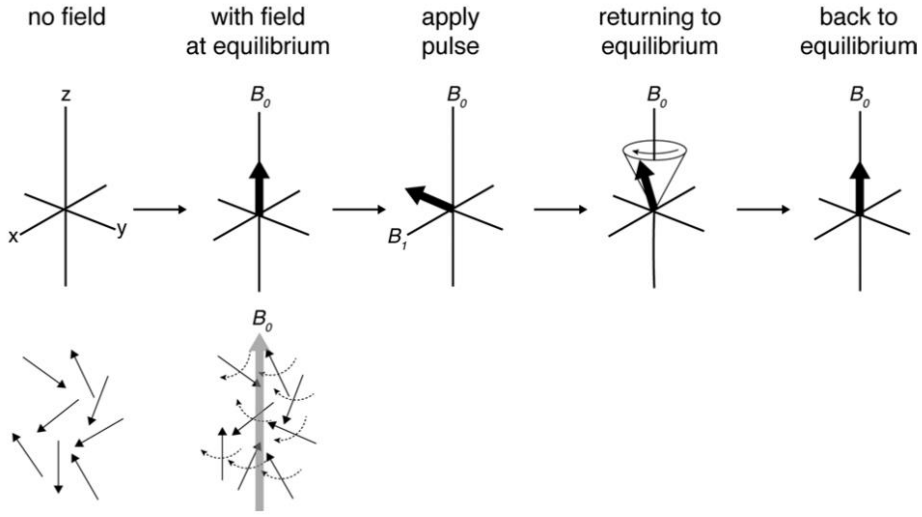


Figure 7: *NMR experiment pulse sequence. Without an external field magnetization vector of each atom are randomly oriented. When an external magnetic field (B_0) is applied, magnetic vectors of atoms sort along the external magnetic field (the z-axis). When short RF-pulse is applied, bulk magnetization is pushed to (x-y plane) and immediately starts to rotate back to equilibrium (along the z-axis) at its Larmor frequency (chemical shift), measured as decaying oscillating electric field. Adapted from*[10].

Compared to X-ray crystallography, NMR beneficially measures samples in solution, enabling the study of protein dynamics in milliseconds time scale, which is not easily accessible by other high-resolution techniques[33]. However, NMR has severe limitations, such as the use of isotopically labelled proteins, higher sample consumption (in order of milligrams) and protein maximal size limitation ~50 kDa for de novo NMR[34].

1.1.2.4 Cryogenic electron microscopy

Cryogenic electron microscopy (Cryo-EM) is the most evolving technique in structural biology these days, overcoming the drawbacks of X-ray (no crystal needed) and NMR (molecule size limit). Its ground-breaking series of discoveries date back to the 70s-80s when crucial sample preparation (vitrification) and data processing (combination of 2D tomograms into the 3D structure) protocols were published. For Cryo-EM development,

Jacques Dubochet, Joachim Frank and Richard Henderson were awarded by the Nobel Prize in 2017, pointing to the importance of this technique[35].

In general, Cryo-EM shares its basic principles with transmission electron microscopy (TEM), where the beam of electrons of similar wavelengths as atoms (~1 Å) interact with the thin layer of the sample. Transmitted electrons are carrying information about the sample, which is then transformed into an image. However, electron microscopes operate under a high vacuum with high energy electron beams, causing proteins burning and evaporation of water molecules stabilizing 3D protein structure[10]. Hence, Cryo-EM works within cryogenic temperatures where the sample is prepared by a process known as vitrification. During the sample preparation, the aqueous sample is flash-frozen in liquid ethane, resulting in the formation of amorphous ice surrounding protein molecules in randomly frozen orientations with preserved native 3D structure[36]. Subsequently, multiple 2D images (thousands of micrographs) are recorded with protein molecules in all possible orientations (**Figure 8**), and sophisticated image-processing software is used to classify groups of "similar" projections. Later on, groups (known as classes) of different orientations are combined into a low-resolution 3D structure or map, and each particle is then fitted to this initial 3D model. By fitting each particle and re-aligning and re-projecting, the enhanced high-resolution model is obtained[37]. To obtain the structural model of the atomic resolution, the polypeptide chain of the known sequence is manually fitted to the calculated model and corrected according to basic geometric parameters (like in X-ray crystallography)[38].

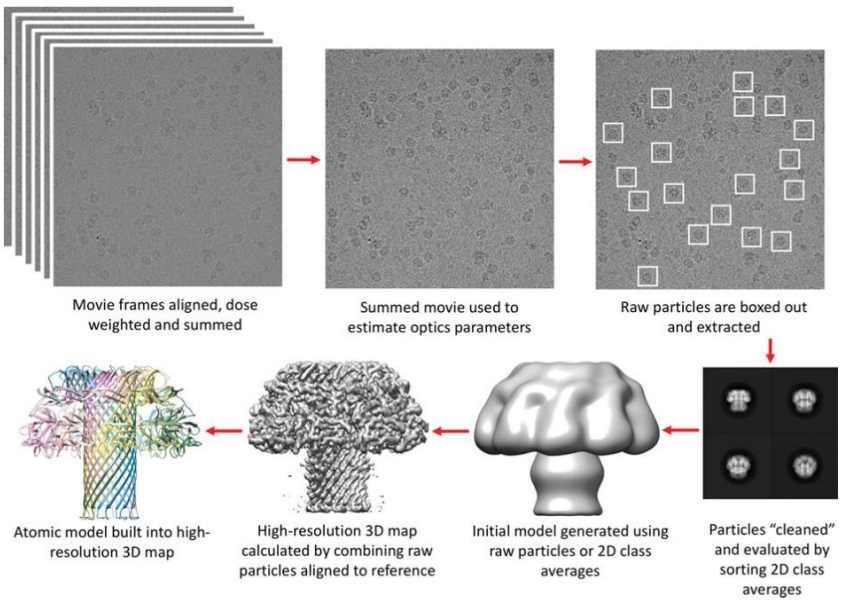


Figure 8: *Cryo-EM image processing outline of pore-forming lysenin toxin. Adapted from*[38].

Since the first publication, the invention of Cryo-EM revolutionized the structural biology field, revealing structures of large macromolecular assemblies not achievable by any known technique. Nowadays, Cryo-EM is pushing limits forwards by enhancing instrumentation which becoming capable of insight into molecular structures at an atomic resolution[39].

1.2 Structural mass spectrometry

Up to the 1980s, mass spectrometry was primarily focused on analyzing *m*/*z* of small molecules. However, in the late 1980s, pioneering publications evolving soft ionization techniques were published introducing electrospray ionization (ESI)[40] and matrix-assisted laser desorption/ionization (MALDI)[41–43]. These key publications opened brand new possibilities enabling the transfer of large molecules into the gas phase without significant fragmentation during the ionization process[44]. Since then, mass spectrometry has made a giant step forward, enabling a detailed study of biomolecules, including their higher-order structure and dynamics in combination with covalent labelling, chemical cross-linking, native mass spectrometry and ion mobility (**Figure 9**)[45,46].

Although structural MS is a potent tool in structural biology, it provides only "low resolution" information and, thus combination with established biophysical methods such as X-ray crystallography, nuclear magnetic resonance or cryo-electron microscopy is beneficial or even essential. However, mass spectrometry occupies a special place compared to the mentioned methods due to its simplicity (like the absence of crystalization step in XRC) and very low sample consumption[47].

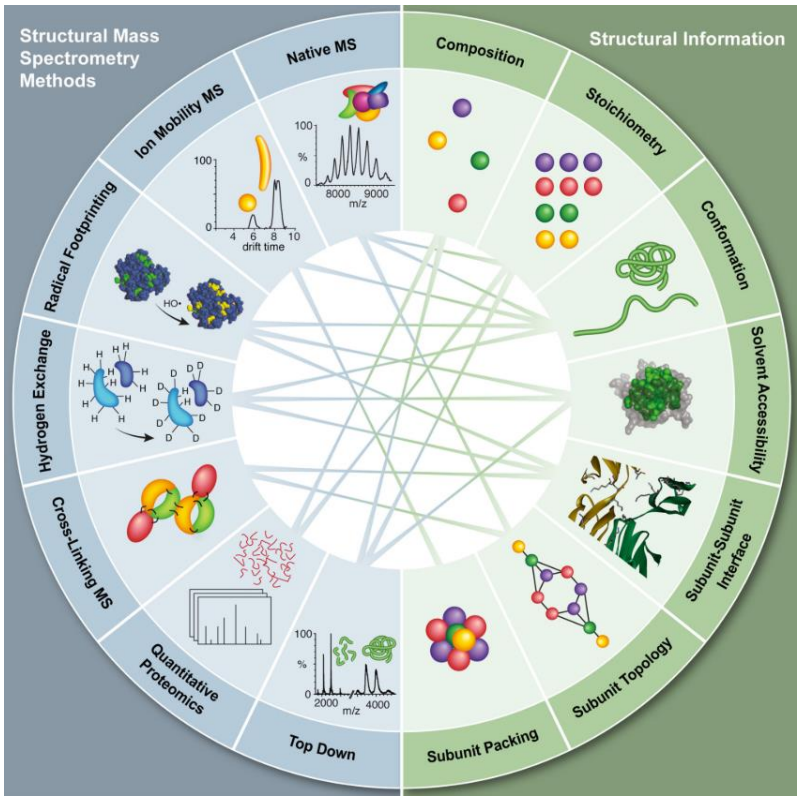


***Figure 9**: An overview of standard structural MS techniques linked with the information they can provide. Adapted from*[45].

1.2.1 Native mass spectrometry and Ion mobility

The invention of ESI has led to a significant boom in the field of mass spectrometry in terms of the development of methods for the study of biomolecules. One of the newly established methods was native mass spectrometry. In 1991, the first publications showed the electrospray's ability to ionize and transport intact protein complexes to the gas phase, defining a milestone of native MS[48,49]. But surprisingly, the name "native" MS was formally introduced more than one decade after its first use in 2004 by Heuvel and Heck[50].

To date, native MS serves as a complementary technique to the portfolio of structural biophysical methods[51]. Native MS versatility was proven by many publications where stoichiometry, topology or kinetics of small proteins to huge protein machineries (like the ribosome, urease, or viral capsids) were solved (**Figure 10**)[52–54].

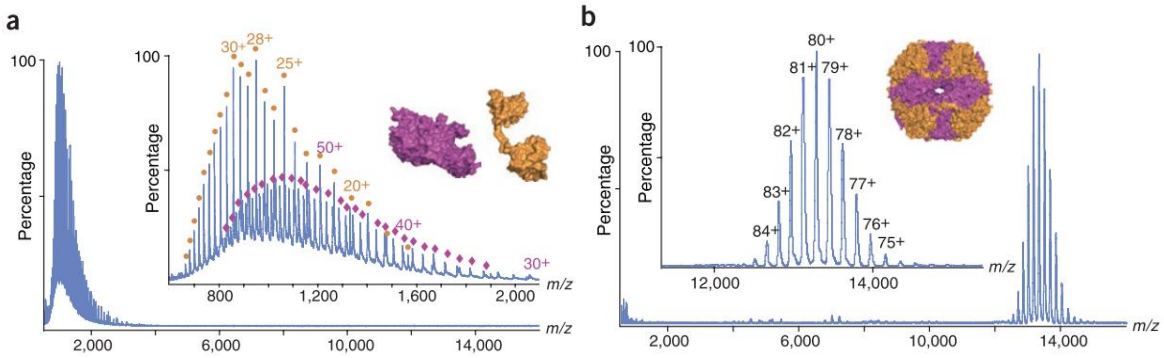


Figure 10: *Comparison of of H. pylori urease ESI spectra measured under (a) denaturing conditions (50:50 (v/v) AcN:H_2O with 0.1 % (v/v) formic acid) revealing distinct envelopes of α (26.6 kDa, orange) and β (61.7 kDa, magenta) subunits. In comparison (b) native conditions (200 mM ammonium acetate, pH 6.7) preserved fully intact dodecameric $\alpha_{12}\beta_{12}$ machinery of total mass 1,063.4 ± 1.0 kDa. Adapted from*[46].

Nevertheless, the commonly used term "native" MS could confuse scientists outside of the mass spectrometric community. The perplexity of the term is originating from biological fields where it is related to the state inside the cells (pH, ionic strength, protein crowding, etc.)[55,56]. However, in-cell "native" conditions cannot be achieved during MS analysis because most mass spectrometers work under low pressures. Thus, native in mass spectrometry is related to the state of a sample prior to the ionization.

Unlike in standard ESI, native MS requires mimicking native-like conditions by using non-denaturing aqueous solvents containing broad concentration ranges of volatile salts (e.g. ammonium acetate, ammonium carbonate). Also, precise pH and ionic strength control play an essential role in native MS[57,58]. However, concentrated salt buffers within traditional ESI sources are problematic, leading to protein-salt adducts formation, making the mass spectra almost indecipherable. In 1996, Wilm and Mann[59] partially solved this problem by miniaturizing the ESI tip diameter to ~1 µm (nanoESI), which made it more tolerant towards salt adducts formation[60,61]. Additionally, a potential of only 0.5-1.5 kV was found to be sufficient to maintain steady spray without the need for external spraying force. Also, flow rates were drastically reduced to only a few nanoliters per minute. In summary, all mentioned benefits make native MS utilizing nanoESI a perfect combination for techniques such as ion mobility, where nativity can be determined through ions collision cross-sections (CCS).

Ion mobility, also called ion mobility spectrometry (IM or IMS), is a method based on the separation of ions in the gas phase in an electric field. The separation of ions of different size to charge ratio occurs through collisions with inert buffer gas, which is introduced into

the drift tube in the opposite direction towards ions. Due to the different ions collision cross-sections, larger molecules undergo more collisions with inert gas than small ones. The separation of ions is then recorded on a detector plate as ion current over time (in tens to hundreds ms scale), giving ion mobility spectrum. This basic principle is summarized in **Figure 11**.

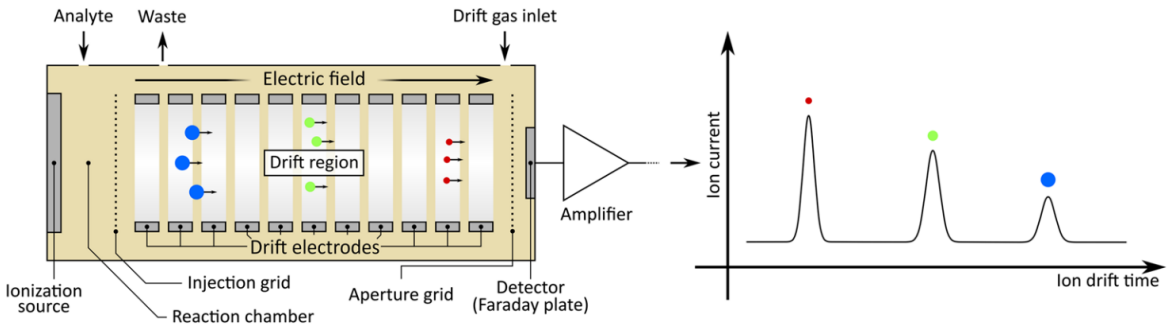


Figure 11: *Basic principle of ion mobility separation in the drift tube IMS. Adapted from*[62].

From the physical perspective, the separation of ions in the limiting conditions (homogeneous gas, weak and uniform electric field ($E = \Delta V/L$)) can be described[63]. Under these conditions, ion's mobility K is constant and is defined as the ratio between steady-state net ion/gas relative velocity $v_d = L/t_d$ and applied electric field E.

$$K = \frac{v_d}{E}$$

However, K depends on numerous variables such as temperature, pressure, polarizability and composition of drift gas etc. Thus, normalization to standard temperature T_0 (273.15 K) and pressure p_0 (101 325 Pa) is often used to obtain reduced mobility K_0.

$$K_0 = K\frac{T_0}{T}\frac{p}{p_0}$$

Reduced ion mobility is valuable information for stand-alone IMS devices, as these values are included in compound libraries for fast and straightforward identification (e.g. airport screening, food quality control)[64].

Nevertheless, the first hyphenation of ion mobility spectrometry-mass spectrometry (IMS-MS) in 1970 by Cohen and Karasek[65] empowered the creation of an even more powerful tool due to the parallel acquisition of ion mobility and m/z in one experiment. This invention gives MS an additional separation dimension which is beneficially used nowadays in proteomics, lipidomics etc.[66,67]. However, obtained m/z and K_0 are satisfactory

parameters to determine ions orientationally-averaged CCS by applying the Mason-Schamp equation[68].

$$\Omega(\mathrm{CCS}) = \frac{3}{16}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{ze}{NK}$$

Where μ is the reduced mass of the ion-gas molecules ($\mu = m_i m_g/(m_i+m_g)$) m_i and m_g are the ion and gas particle masses, respectively. k_B is Boltzmann's constant, T is the drift gas temperature, z is the nominal ion charge, e is the elementary charge, N is gas number density, and K represents ion mobility.

Despite that, calculated CCS must be evaluated with caution because the Masson-Schamp equation is derived from a simplified mathematical model (but the best we know yet) approximating zero electric field within the IMS experiment, which is not feasible during the real experiment[69]. Thus, the primary output of IMS is *K,* with limitations that were already described in the previous text.

Nonetheless, described equations are solely valid for the most straightforward instrument setup, Drift tube IMS (DTIMS). Its aspects were already described at the beginning of this chapter and can be seen in **Figure 11**. Nowadays, instruments with different separation principles capable of IMS-MS are commercially available (**Figure 12**). However, most of these instruments cannot provide CCSs values directly due to the non-homogeneous electric field used for ions separation. Thus, DTIMS still serves as the primary method for determining direct CCSs even though the capability of differential mobility analyzers (DMA or DIMS) measuring directly CCS was also proved.

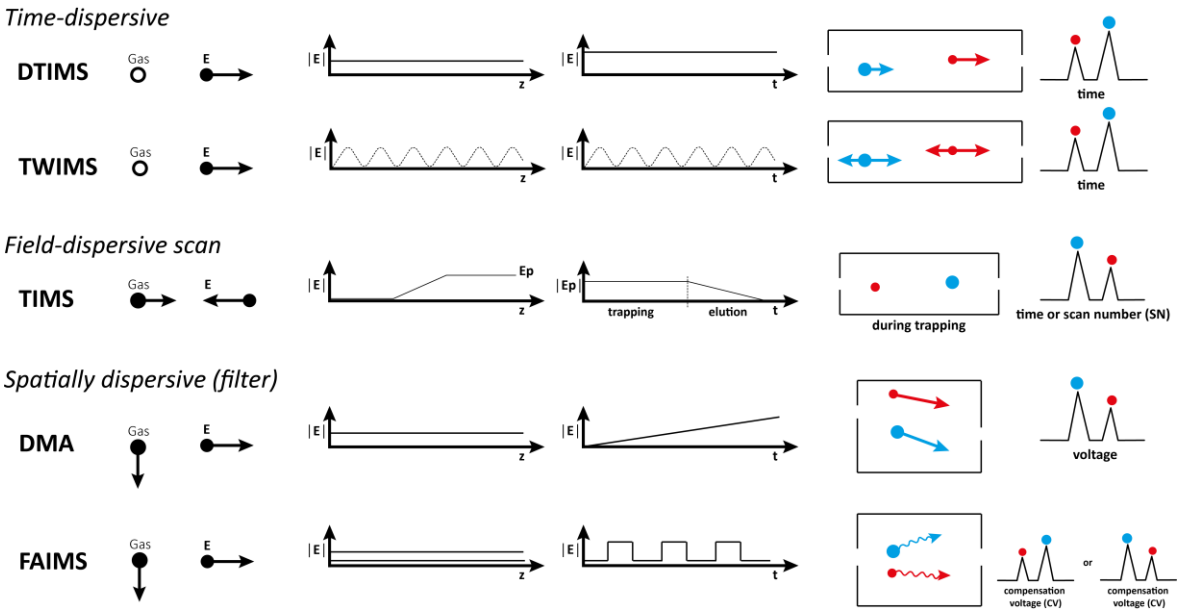


***Figure 12**: Collection of common separation principles used in ion mobility; DTIMS = drift tube IMS; TWIMS = travelling wave IMS; TIMS = trapped IMS; DMA =differential mobility analyzer; FAIMS = field-asymmetric waveform IMS. From left to right: acronym, gas direction, electric field direction, the axial profile of electric, time profile of an electric field, ions motion and separation readout. Edited and adapted from Gabelica et al.*[70].

However, the most widespread technique for the IMS-MS experiment is ion mobility with a travelling wave (TWIMS). In this case, ions separation occurs within static drift gas filled in stacked ring ion guide electrodes to which series of RF voltage is applied to form potential travelling wave propagating towards MS detector. These pulses are propelling ions through drift gas, separating ions by their size (large ions are delay through more collision than small ions). By alternating travelling wave speed and its voltage magnitude separation can be fine-tuned. Although TWIMS is operating below the low-field limit as in DTIMS, direct CCS cannot be easily assessed due to the constantly alternating electric field resulting in the invalidity of the relationship between CCS and K[70,71]. Hence, external calibration by molecules of known CCS (standards) is inevitable[72]. To achieve optimal calibration, standards of similar chemical and physical properties should be used. Fortunately, many works have been published providing directly measured CCS values for small molecules[73], peptide polymers[74], denatured proteins, native-like proteins[75], oligonucleotides[76] and N-glycans[77].

Since the first commercially available IMS-MS instrument was Synapt (Waters) equipped with TWIMS[78] (in 2006), this instrument ignited a revolution in the structural MS by

facilitating CCSs and dynamics of large protein complexes like GroEL-GroES[79] or virus assembly intermediates[80].

Another less common separation principle is the High-field asymmetric waveform IMS (FAIMS). In this setup, ions are introduced between two electrodes where the periodic asymmetric waveform is introduced. Separation occurs due to the different mobilities at high (>7500 V cm^{-1}) and low electric field[81,82]. This principle has been advantageously used within proteomics experiments, where IMS separation filtered out abundant precursor ions allowing the instrument to analyze even low abundant peptides within the limitation of instrument duty cycle[83].

In 2011 new IMS separation named trapped ion mobility spectrometry (TIMS) was introduced[84]. During the TIMS experiment, ions entering the ion guide are trapped by the non-uniform electric field, which holds ions against moving drift gas. Trapped ions are then spatially separated along the axial direction, trapped in the electric field at the position according to their ion mobilities (compensation of drift force vs electric field). Scanning is then performed by the continuous decrease of trapping electric field causing step-wise elution of trapped ions from high to low size-to-charge ratio towards MS detector[85].

In addition to the standard ion separation methods in the previous text, many others exist, such as differential mobility analyzers (DMA)[86], cyclic ion mobility spectrometry (cIMS)[87] etc. However, their practical use is limited to particular applications, and their detailed descriptions are summarised in a review by Cumeras *et al.*[64].

The native IMS-MS becomes an effective tool in structural biology, providing information that other complementary methods cannot answer. However, the native IMS-MS method raised many doubts about its validity as the mass spectrometers operate at low pressures close to vacuum. On the concerns about the "nativity" of biomolecules in the gas phase Gabelica in 2019, noted that in "….*fields like structural chemistry and biophysics, there is always this suspicion about native MS, because we are analyzing things in the gas phase and we want to infer information on what existed in the solution. In every community you have to convince and prove that what you are claiming to deduce from MS is valid. In different communities it's a different challenge. For example, among organic chemists and supramolecular chemists, a few are really convinced, but most will want to crystallize the structure and analyze it by NMR, because that is more conventional and thus better accepted by their peers."*[88]. Even though her opinion on native MS, D'Atri and Gabelica[89] published a manuscript where alterations of G-quadruplex structures were found upon transfer from

solution to the gas phase (structure compared to circular dichroism data). She assigns this phenomenon to the memory of the solution structure in the gas phase due to the presence of native MS buffers manipulating charge states[89]. However, other publication from Borysik *et al.*[90] tested globular proteins of broad mass range (15 kDa-500 kDa), revealing good conformity of measured CCS values with Stokes radii (R_s) obtained from Size-exclusion chromatography coupled with dynamic light scattering (SEC-DLS).

Aside from "nativity", structural IMS-MS can explore biomolecules stability in the gas phase through collision-induced unfolding (CIU). As proof of concept, the work published in 1997 by Shelimov and Jarrold[91] showed progressive structure unfolding of apomyoglobin with a gradual increase of injection energy without protein fragmentation.

In a typical CIU experiment shown in **Figure 13**, molecular ions of interest are isolated and subjected to collisions with inert gas, resulting in increased internal energy of selected ions. Subsequently, collision voltage is ramped in defined steps resulting in a gradual increase of internal energy enabling molecular ions conformation change (predominantly unfolding), which IM detects as the transition of drift time or CCS. After that, drift times across a measured range of collision voltages are plotted, creating the 3D heatmap, also know as CIU fingerprint[92].

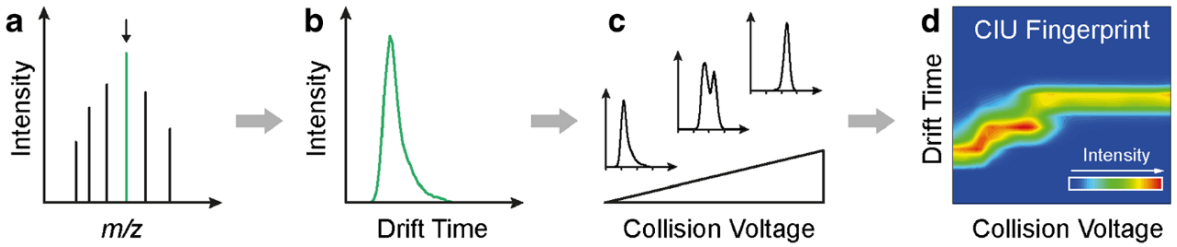


Figure 13: *Typical of CIU experiment workflow. a) In the first step, the ion of interest is isolated b) and its mobility is measured. c) Isolated ions undergo collision activation by a step-wise increase of collision voltage. d) Finally, measured mobilities are plotted to collision voltage forming CIU heatmap. Adopted from*[93].

For example, the utility of the CIU approach was demonstrated in the work of Tian *et al.*[94] where CIU of 4 different intact antibodies was performed (**Figure 14**). Interestingly, authors distinguished iso-cross-sectional species by unique unfolding pathways related to sums and patterns of disulfide bonds. Since the experiment is conducted in the gas phase by isolation and perturbation of charge state of interest, such dataset can be acquired rapidly compared to in-solution experiments where stability is studied by increasing solution temperature.

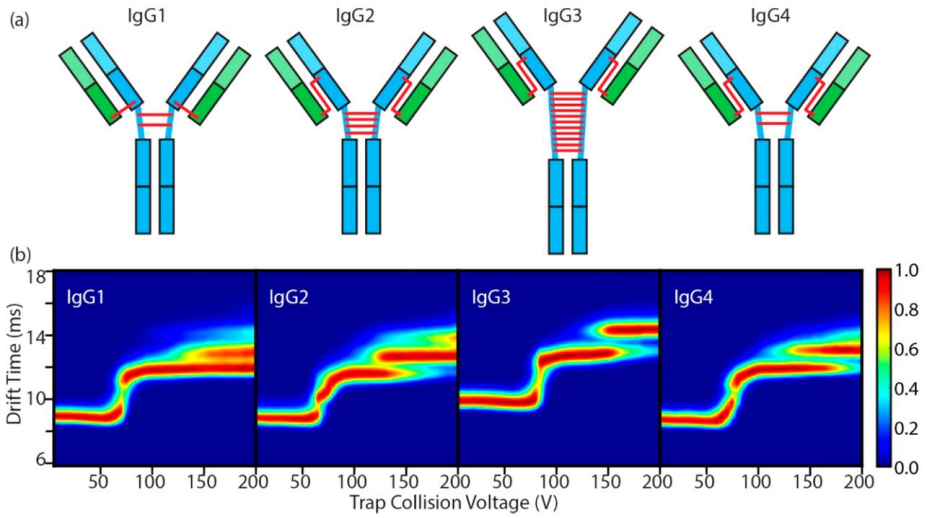


Figure 14: *a) Four different IgG antibodies used for CIU activation. b) Heatmaps are representing CIU pathways revealing differences in pattern and number of disulfide bonds. Adapted from*[94].

As was described in the previous paragraph, in-solution temperature-based stability studies are considerably more time-consuming than CIU experiment. In addition, special equipment like commercially not available variable-temperature electrospray ion source (vT-ESI) is needed to perform such MS experiment. However, the vT-ESI source beneficially allows unique readouts of notable structural changes (system dynamics) during protein thermal denaturation or DNA/RNA solution melting[95,96]. Nevertheless, due to the in-house development of the vT-ESI source, a limited number of research groups presented its practical use yet.

Recently, Woodall *et al.*[97] published work where increased temperature caused the subsequent denaturation of myohemerythrin followed by the release of diiron oxo [Fe(μ-O)Fe] cofactor (**Figure 15**). Interestingly, the formation of non-native disulfide bond was reported (within 83-90 °C). Non-native disulfide formation led to protein stabilization detected by the dramatic decrease of CCS and shift of charge state distribution towards lower values. As Can be seen, such a unique vT-ESI-IMS-MS combination enables novel applications of tracking protein-ligand dynamics, not achievable by any currently used method in structural biology.

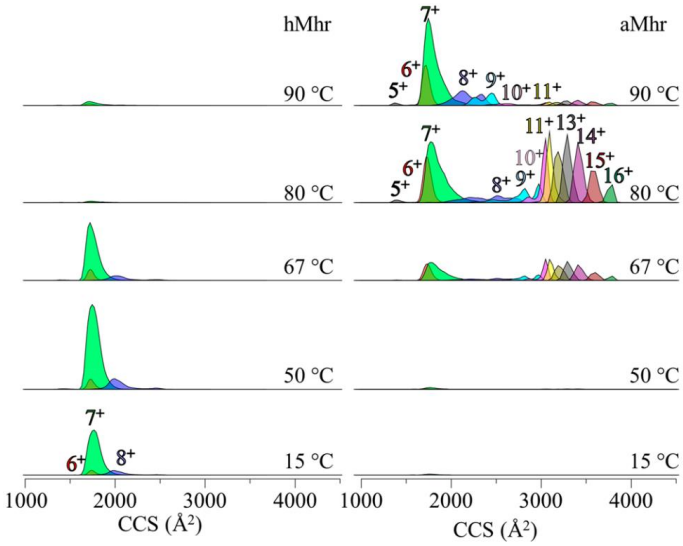


***Figure 15**: Temperature-dependent mobiliograms of detected charge states of holomyohemerythrin (hMhr) and apomyohemerythrin (aMhr). At the temperature of 67 °C, higher charge states and increased CCS of aMhr revealed evidence of protein unfolding. However, at 90 °C sudden shift towards lower CCS and charge states was observed due to the formation of stabilizing non-native disulfide bond. Adapted from Woodall et al.*[97].

1.2.2 Covalent labelling

One of the standard methods of structural MS is covalent labelling (also known as footprinting), which provides important information about the dynamics and HOS of biomolecules. This method relies on reversibly or irreversibly linked probes that incorporate into the protein sequence (structure), which results in a change of biomolecule mass, making them easy to identify and quantify by MS, either by bottom-up or top-down[98].

The term "footprinting" was first used to study DNA-protein interactions where bound protein protected DNA from DNase activity, leaving a footprint of its binding site (**Figure 16a**)[99,100]. In 1988 Payne and co-workers[101] inspired by previous works, introduced the term "protein footprinting", summing up all names used for covalent labelling techniques for HOS determination. In the case of proteins, the footprint is made by reactive probes targeting amino acids side chains exposed to the solvent. Such information provides an evidence of probe reactivity, selectivity and solvent accessible surface area (SASA) of protein or protein complex (**Figure 16b**)[102].

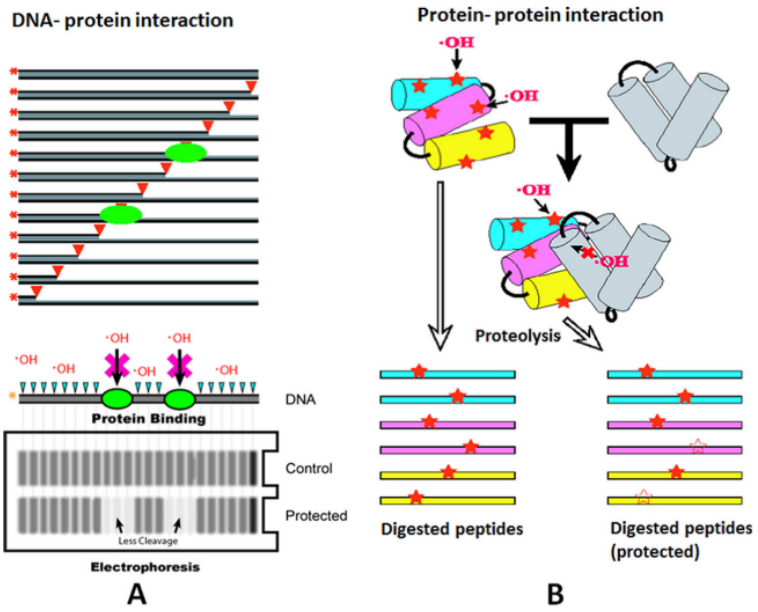


***Figure 16**: Footprinting strategies for mapping a) DNA-protein interaction and b) protein-protein interaction via covalent labelling. a) Green ovals represent protein bounds to DNA. Compared to a controlled experiment, an interrupted DNA fragments ladder is present, revealing DNA protection sides. b) Protein-protein interaction is revealed upon protease cleavage at peptides level where an absence or lower abundance of modifications revealing interaction interphase of the protein complex. Adapted from*[103].

An illustration of the bottom-up footprinting experiment, **Figure 17** summarizes essential steps. The labelling experiment starts by protein dissolution in a compatible buffer. Subsequently, a reactive probe is introduced at the defined ratio over the protein concentration for a well-defined period of time. Whether photolabelling agents are used, UV light irradiation (powerful light lamp or laser) is required to initiate a labelling reaction. The irradiation can be conducted in continuous or pulse mode. The labelling reaction is quenched, and samples at different time points are collected in order to achieve information about the protein dynamics. After that, the labelled samples and subjected to bottom-up or top-down mass spectrometry analysis where the modified residues are identified and quantified[104].

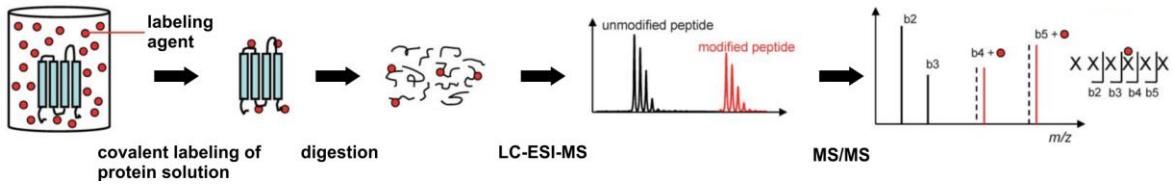


***Figure 17:** Standard labelling experiment workflow. Edited and adapted from*[104].

For structural covalent labelling MS, two separate biological states related to structural changes such as protein folding/unfolding or complex formation/dissociation can be tracked. Identified modifications extents are then reflecting reactivity, structure and solvent accessibility of labelling agent.

The most common way to determine modification extend is to compare relative ion abundances of peptides containing reactive residue before and after the structural change. To calculate the ratio of the modification for a single peptide's fragment ion, the following equation can be used:

$$Modifcation\ extent = \frac{\sum I_{modified}}{\sum (I_{modified} + I_{unmodified})}$$

This ratio is then calculated for both states (folded/unfolded), and by simple comparison, change of modification extent represent the structural event. For example, increased modification extent for protein labelled in folded and unfolded states means that a particular residue (or region) was more accessible toward the labelling probe, probably caused by such a region's unfolding[105,106].

The previous text described the most commonly used bottom-up approach. Despite that, the top-down approach was also utilized. Unlike in bottom-up, the labelled intact protein is directly submitted to the mass spectrometer and being extensively fragmented by a broad pallet of fragmentation techniques[107,108]. Top-down key benefits arise from the capability of selectively fragment protein isoforms (protein without any modification, singly oxidized protein, etc.). However, large proteins yield poor fragmentation providing low spatial resolution. This limitation is also the reason why top-down is predominantly used for smaller proteins[109]. Nevertheless, mentioned drawbacks were partially overcome by the use of less common fragmentation techniques like electron capture dissociation (ECD)[110], electron transfer dissociation (ETD)[111], surface-induced dissociation (SID)[112], ultraviolet photodissociation (UVPD)[113,114], and infrared multiphoton dissociation (IRMPD)[115] predominantly performed with high resolution Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers[116].

The described approaches are performed with irreversibly bounded labelling probes that track the surface through reactions with amino acid side chains. However, structural information can also be obtained by HDX, which is the covalent technique utilizing a reversible probe (hydrogen, deuterium). Additionally, HDX targets the protein backbone,

providing complementary information to previously described covalent labelling techniques[117].

1.2.2.1 Hydrogen deuterium exchange

Hydrogen deuterium exchange is a covalent labelling technique utilizing the reversible exchange of labile hydrogen (^1H = 1.007825 Da) in N-H, O-H or S-H bonds for its heavier analogue deuterium (^2H = 2.014102 Da), or vice versa[8]. Once a protein is dissolved in the deuterated solvent, the exchange is taking place across all exchangeable (solvent-accessible) hydrogens in the protein backbone and amino acids side chains. Nevertheless, exchange/back exchange on amino acid side chains is extremely fast, which makes it very difficult to track. Due to that, just a few publications which perform such experiments were published utilizing HDX exchange in gas phase with ND_3 in the time scale of (0.1−10 ms)[118].

To this day, methods such as UV spectroscopy[119], IR spectroscopy[120,121], neutron diffraction[122], NMR[123–125] and mass spectrometry[126,127] were used to study the HDX. However, these days mass spectrometry is the method of choice due to its low sample consumption, simple sample preparation and the ability to analyze large proteins and protein complexes compared to NMR. Additionally, detection in mass spectrometry is directly connected to mass; thus, HDX (H→ D = +1.006277 Da) is readily detectable. Even though the mass spectrometric approach is dated back to the 1990s where the first experiments were conducted, fundamentals of HDX were investigated back to the 1950s.

In 1954 the basic principle of hydrogen-deuterium exchange was firstly proposed by Linderstrøm-Lang and Hvidt[8,128]. In their work, they used purified pork insulin placed in 99.6% deuterium oxide. In desired times, they were stopping the deuterium exchange by freezing the samples to -80 °C, followed by lyophilization and drying in a vacuum for several days. After that, they determine the amount of exchanged deuterium by the dissolution of dried samples in pure water and by comparison based on the densities of pure and deuterated water. Surprisingly, they found only 46 hydrogen atoms exchanged out of a possible 92, even after an extended exchange time of 124 hours. This observation led them to the hypothesis that resting hydrogen atoms are involved in more stable N−H···O=C bonds within the protein backbone. Obtained results were also in agreement with the work of Pauling L. and Corey R. B.[5] where the protein secondary structure stabilized by hydrogen bonds was proposed. In 1966 Hvidt and Nielsen[129] published a comprehensive review about

HDX. This work includes a description of exchange kinetics based on the assumption that the exchangeable backbone sites fluctuate between "close" and "open" states described by the rate constants k_{op} and k_{cl}. When the open state is reached, accessible hydrogens can be exchanged for deuterium. This reversible event is described by rate constant k_{ch} in the following equation.

$$\mathrm{N-H_{closed}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} \mathrm{N-H_{open}} \underset{D_2O}{\overset{k_{ch}}{\rightleftarrows}} \mathrm{N-D_{open}} \underset{k_{op}}{\overset{k_{cl}}{\rightleftarrows}} \mathrm{N-D_{closed}}$$

This equitation is valid for every N-H group on the peptide backbone with unique values of all rate constants. Based on the values of rate constants, two limiting regimes known as EX1 and EX2 kinetics can occur (**Figure 18**). In EX1 kinetics ($k_{ch} \gg k_{cl}$), amide exchange occurs under a single or a few opening/closing cycles. When opening and reclosing is much faster than HDX ($k_{ch} \ll k_{cl}$), the probability of deuteration upon a single opening/closing cycle is low. To achieve successful deuteration, multiple of these events must occur. This phenomenon is known as EX2 kinetics and can be easily observed by mass spectrometry as a continuous shift of isotopic pattern to higher *m*/*z*. On the contrary, EX1 kinetics can be distinguished due to the presence of a separate population in mass spectra[130]. In naturally occurring proteins, EX2 kinetics was found to be predominant; however, the combination of EX1 and EX2 known as EXX is also common[131–133].

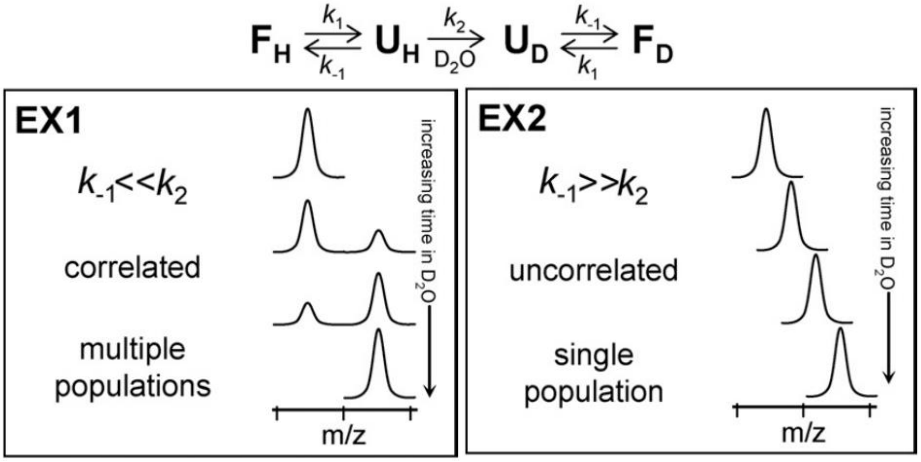


Figure 18: *Illustration of EX1 and EX2 deuteration kinetics of protein sample observed by mass spectrometry. Adapted from*[130].

The next crucial factor affecting HDX is the pH of the sample environment alternating exchange rate by the phenomenon known as acidic or basic catalysis. In addition, water catalysis was also described, but its effect is insignificant[134]. When the dependency of rate

constant for averaged amino acid on pH is plotted, a distinctive "V-like" curve with a minimum around 2.5-3.0 pH is obtained (**Figure 19**)[135–137]. At this point, the exchange rate is the lowest. This phenomenon is advantageously used in the HDX experiments primarily conducted under native-like conditions (pH ~7.0), where rapid pH change to 2.5-3.0 significantly decreases the sample's exchange/back-exchange rate (quenching of HDX)[137].

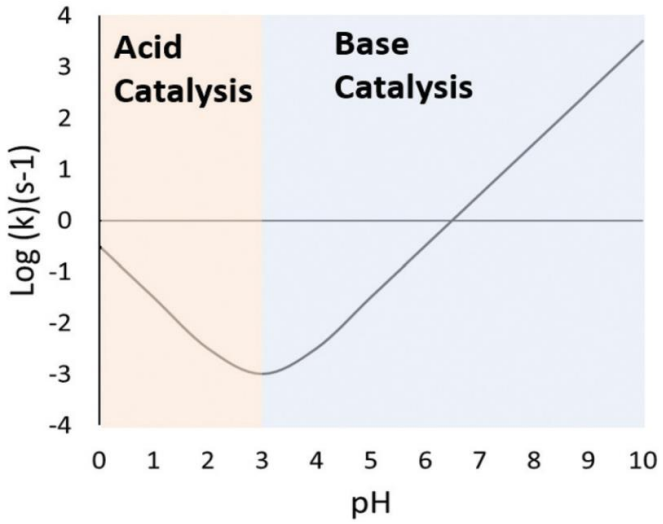


Figure 19: *Dependence of exchange rate on the solution pH. The minimum value around pH 2.5-3.0 highlights the point where the exchange/back-exchange rate is minimal. Adopted from*[137].

Last but not least, the temperature is another parameter influencing HDX. Its effect originates from the fundaments of kinetics postulated by Arrhenius in 1889[138], which define that a change of 10°C will result in a 2.8x decrease or increase of reaction speed[139]. In HDX, a modified Arrhenius equation is used[136].

$$k_{ch}(T) = k_{293} \exp\left(-\frac{E_a}{R}\left[\frac{1}{T} - \frac{1}{293}\right]\right)$$

Where k_{293} is the rate constant for acid or base or water-catalyzed reaction at 20°C, the E_a is the corresponding activation energy (14, 17 and 19 kcal mol^{-1} for acid, base or water-catalyzed reaction, respectively). R is the gas constant 8.314 J mol^{-1} K^{-1}.

Consequently, a rapid decrease of the temperature (e.g. flash freezing in liquid nitrogen) causes a significant drop in the exchange rate, partially preventing additional deuteration or loss of deuteration caused by the back-exchange. Due to that, the HDX experiments are conducted with instrument parts cooled to minimize this effect (mainly back-exchange)[140].

To perform the HDX experiment, two different approaches were developed. The first and most widely used is continuous labelling (**Figure 20a**) to study different protein states such as unfolded/folded protein or free protein/protein-ligand[141]. For this experiment, the sample

is diluted to 90% D_2O buffer, and after the desired time periods (in the range of seconds to days), part of the sample is quenched by diluting in the buffer of pH around 2.5 and flash-freezing. Thus, continuous labelling experiments are used to study the HDX kinetics of the amide backbone. On the contrary, pulsed labelling (**Figure 20b**) aims to process kinetics like structural changes in time (e.g. protein aggregation[142], pH-induced unfolding[143]). Therefore, the labelling time (HDX) and quenching remain constant while the studied system is perturbated in time.

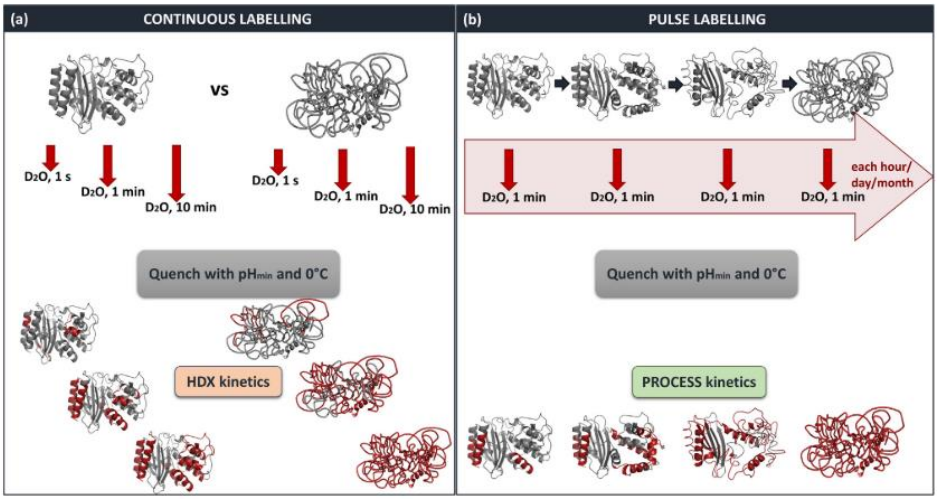


Figure 20: *Comparison of a) continuous labelling and b) pulse labelling HDX experiment workflow. Adapted from*[144].

To pinpoint exchanged regions of a protein, the top-down or bottom-up approach can be used. The top-down approach employs the labelled sample directly to the instrument, while the bottom-up approach has extra steps as protein digestion and separation by LC-MS. Despite the extra steps, the bottom-up approach becomes the leading in HDX-MS analyses[145].

In the case of highly dynamic proteins (e.g. intrinsically disordered proteins or proteins with partially unstructured regions), structural events are so fast that standard HDX sample preparation cannot track them. In such cases, quench-flow apparatuses composed of quartz capillaries and small mixers can achieve labelling time in the range of tens of milliseconds, capable of freezing fast transitional proteins states[146–149].

1.2.2.2 Chemical labelling

Chemical labelling of biomolecules, also known as bioconjugation, is an enormous set of techniques used in broad applications like medical, diagnostics, life science, microelectronics, materials, etc. In its nature, the process of bioconjugation is understood as the attachment of biomolecules to other molecules, mainly through a covalent bond. The bioconjugates can be divided by their application into six fundamental groups: (1) assays and quantification; (2) detection, tracking and imaging; (3) purification, capture and scavenging; (4) chemical modification; (5) *in vivo* diagnostics and therapeutics; (6) vaccines and immunity modulation[150].

To successfully form bioconjugate, sufficient knowledge of targeting groups reactivity and occupancy, reagent types, and structural aspects must be kept in mind. However, due to the variedness of biomolecules and reactivity properties of diverse reacting groups, this chapter will be focusing only on proteins, especially on the reactivity of specific amino acids groups and their applications for structural MS[151].

The first chemical modification of amino acid can be traced back to 1904 where the modification of histidine was described by Pauly[152]. Until now, many protein applications were utilized in bioconjugation, such as in-cell localization through attached fluorescence probe[153]; identification of interacting partners by chemical cross-linking[154]; *in vivo* biodistributions via radiolabels[155] or detection by immobilized immunoassay used in clinical laboratories[156].

For structural purpose, protein labelling experiments are typically conducted under pH between 6 and 8.5. Under these conditions, protein backbone and amino acid side chains are generally stable, and their HOS is preserved. Structural MS experiments are usually performed at temperatures below 37 °C and with minimal concentration of organic solvents (DMSO, DMF or methanol), which are often used for reagent solubilization.

Due to the protein native fold, amino acids whose polar side chains are exposed to the solvent are labelled predominantly. The frequency of amino acid occurrence on the protein surface is summarized in **Figure 21** and reveals the most exposed polar amino acids whose side side chains (primary amines, carboxyls, etc.) are suitably available for a labelling reactions. Therefore, a substantial part of the following paragraphs will be devoted to reactions aimed at modification of these groups, especially on reactions with of primary amines.

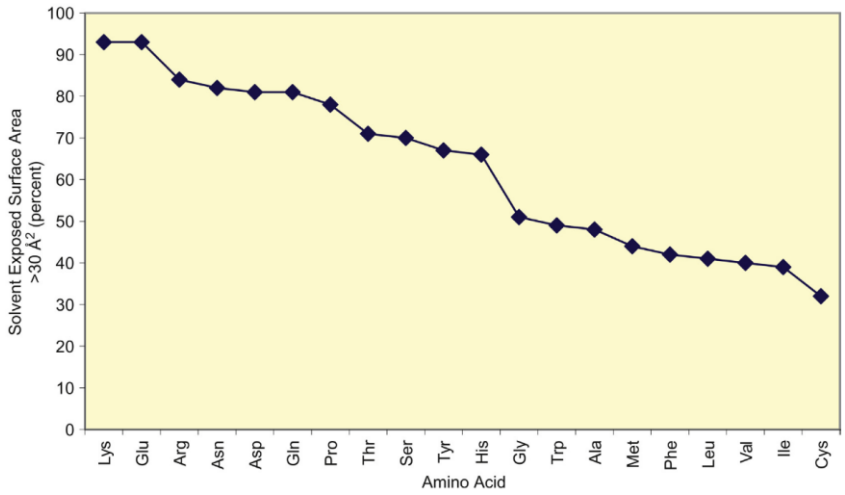


Figure 21: *Comparison of the solvent-exposed surface area of amino acids in proteins data represents the percentage of each amino acid in the protein having an exposure to the aqueous solvent greater than 30 Å². In general, charged and polar amino acids are significantly more exposed to solvents than uncharged, aromatic or aliphatic amino acids. Adapted from*[150].

As was written, lysines primary amines (also N-terminal amines) serve as the gold standard for bioconjugation. Even under physiological pH ~7.4, when lysine ε-amine are protonated (pK_a ~9.2[157]), they still react as nucleophiles. The most common labelling chemistry is based on N-hydroxysuccimide (NHS) esters (**Figure 22a**). The reaction of NHS with primary amine leads to the formation of an irreversible amide bond. The NHS derivates are commercially available for a broad range of application (e.g. introduction of fluorescence probe or affinity tags)[150]. However, NHS esters are poorly water-soluble and require organic solvents like DMSO for dissolution. The synthesis of water-soluble sulfo-NHS analogues overcame this issue.

In a similar fashion, isothiocyanates react with primary amines forming thiourea analogues (**Figure 22b**). Nevertheless, optimal reaction pH is 9.0-9.5, leading to unstable modification products in physiological pH.

Nucleophile attack towards 1H-benzo[d][1,3]oxa-zine-2,4-dione (**Figure 22c**), rapidly results in ring-opening, forming orthoaminebenzamide, which could be then oxidatively functionalized[158].

The next reaction (**Figure 22e**) is based on the reductive alkylation by aldehyde, yielding in the formation of a reversible Schiff base which is then reduced to a form of secondary amine[159].

To overcome unstable Schiff base intermediate, 6π-aza electrocyclization was proposed by Tanaka *et al.*(**Figure 22f, g**)[160].

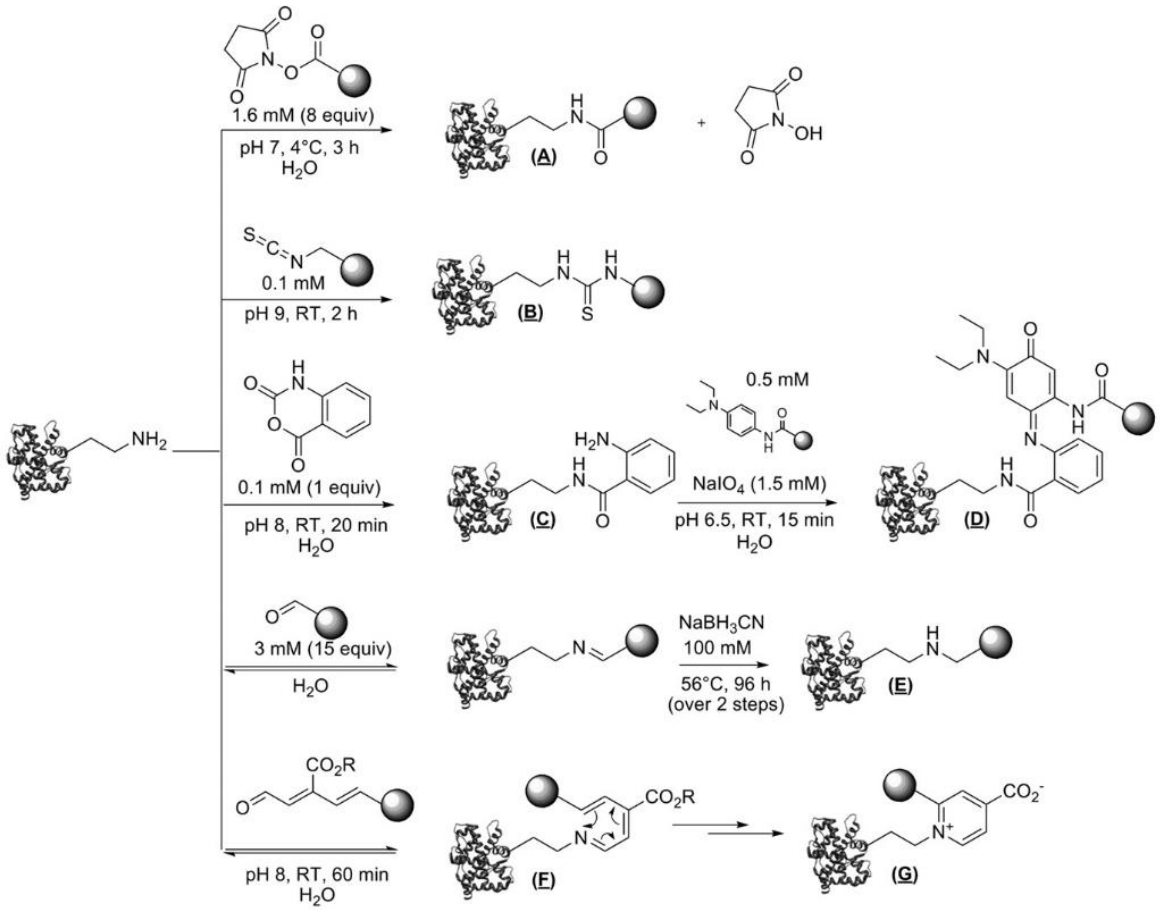


Figure 22: *Chemical strategies for modification of lysine residues or N-terminus. Adapted from*[161].

Another widely used chemical labelling target is cysteine. Compared to lysine, cysteine is a rarely occurring amino acid. On top of that, reduced cysteines are unusually found on the protein surface due to the significant contribution to the protein ternary structure via the disulfide bonds[162].

Cysteine in its reduced state in physiological pH conditions can be deprotonated (pK_a ~8) to form thiolate nucleophile. This cysteine form can undergo alkylation reactions with iodoacetamide (IAA), maleimides or vinylsulfones (**Figure 23a, b, c**). Cysteine alkylation is commonly used as subsequent modification after the reduction of disulfide bridges (**Figure 23e**) by dithiothreitol, mercaptoethanol or tris(2-carboxyethyl)phosphine (TCEP) within proteomics experiment[163].

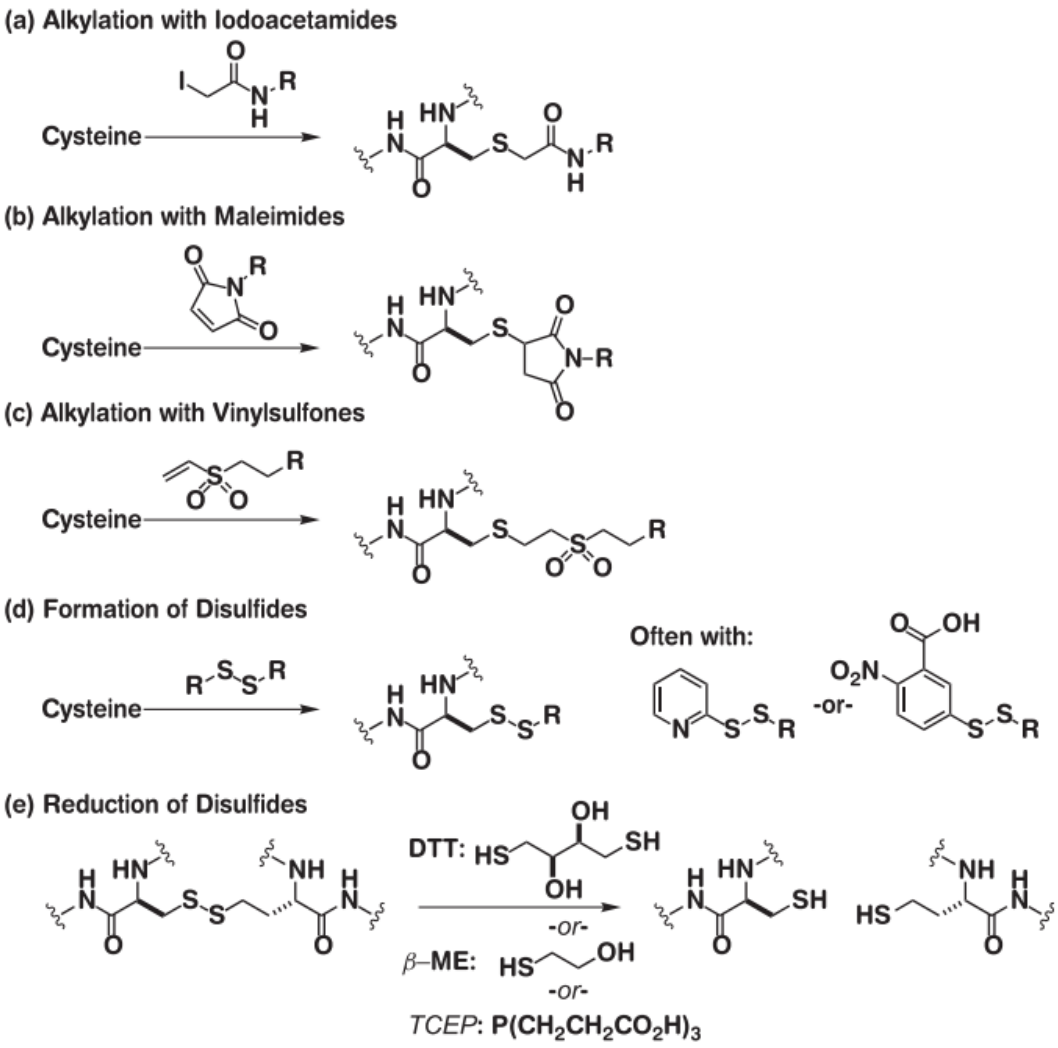


Figure 23: *Chemical strategies for modification of cysteine residues. Adapted from*[164].

Further, glutamic and aspartic acids are polar amino acids predominantly found on the protein surface, exposed to the solvent, making them perfect candidates for chemical labelling. Their carboxyl groups (also C-terminus) can be activated by water-soluble carbodiimides such as *N*-ethyl-3-*N'*,*N'*-dimethylaminopropylcarbodiimide (EDC).

The EDC reaction optimum was found to be in the range of pH 3.5-4.5 in aqueous solutions. However, EDC is less stable at low pH, far away from native conditions. Thus reaction is commonly performed at pH 6.5 and above, where EDC stability is prolonged at the cost of significantly slower reaction[165,166].

The two-step reaction mechanism (**Figure 24**) starts with the formation of reaction intermediate *O*-acylisourea, which can then react with a primary amine (e.g. lysine side chain). This reaction yields an amide bond (pseudo peptide bond), creating a zero-length

cross-link. Besides the cross-link, *O*-acylisourea can turn in isourea byproduct, decreasing the total efficiency of the desired reaction product.

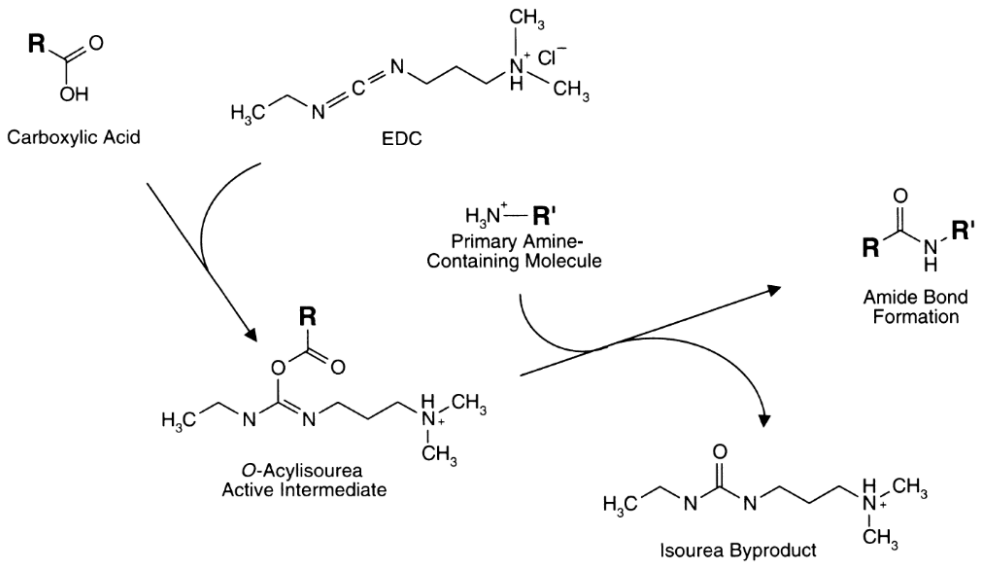


Figure 24: *Chemical modification of carboxyl group presented in aspartic and glutamic acid residues (C-terminus shows similar reactivity) with EDC. Adapted from*[150].

Last but not least, tyrosine labelling approaches are shown. In **Figure 25a,** the diazonium compound reacts with tyrosine through the electrophilic attack of the diazonium group towards electron-rich positions at the aromatic ring. In the case of tyrosine, the electron-rich region is in the *ortho* position to the aromatic hydroxyl. This reaction is forming diazo compounds with characteristic coulours[150].

The following reaction is the iodination of tyrosine (**Figure 25b**) used in biomedical research, especially for the introduction of radiolabelled ^{125}I and ^{131}I[167,168].

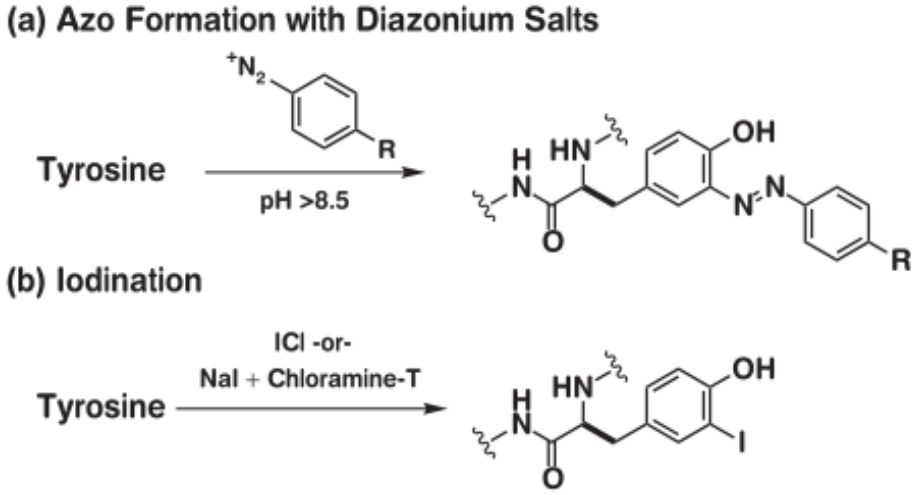


Figure 25: *Chemical strategies for modification of tyrosine residues. Adapted from*[164].

As an alternative approach to conventional bioconjugation techniques, photochemically induced bioconjugation techniques were developed. Unlike previous techniques, where reactions are highly specific, the photolabelling compounds undergo highly reactive states, which makes them reactive toward the molecules in close proximity without defined selectivity. The following paragraph is focusing on the selected photosensitive groups such as aryl azides, benzophenones and diazirines[150].

The first group discussed are the most used aryl azides (**Figure 26a**). Upon UV irradiation (250-400 nm), reactive nitrene is formed and immediately reacts with any C-H or heteroatom-H bond, inserting active hydrogen. The addition reaction with a double bond is also possible. However, ring expansion followed by a reaction with nucleophiles was found to be predominant.

Another photoreactive compound class are benzophenones forming reactive triplet-state intermediate upon UV irradiation (350-365 nm) (**Figure 26b**). An activated triplet state can insert similar bonds as aryl azides. Unlike aryl azides, the triplet state of benzophenones can be reversibly deactivated to the ground state without degradation, allowing multiple activations of unreacted labelling agent.

The last included photolabeling agents are diazirines with the reactive group activated by UV light (350-360 nm), forming reactive carbenes (**Figure 26c**)[169]. This reactive carbene intermediate readily inserts C-H or N-H bonds in a similar fashion as previously described compounds. Advantageously, amino acid analogues containing diazirine groups (photo-leucine, photo-isoleucine and photo-methionine) were developed for protein interaction studies within the cells[170].

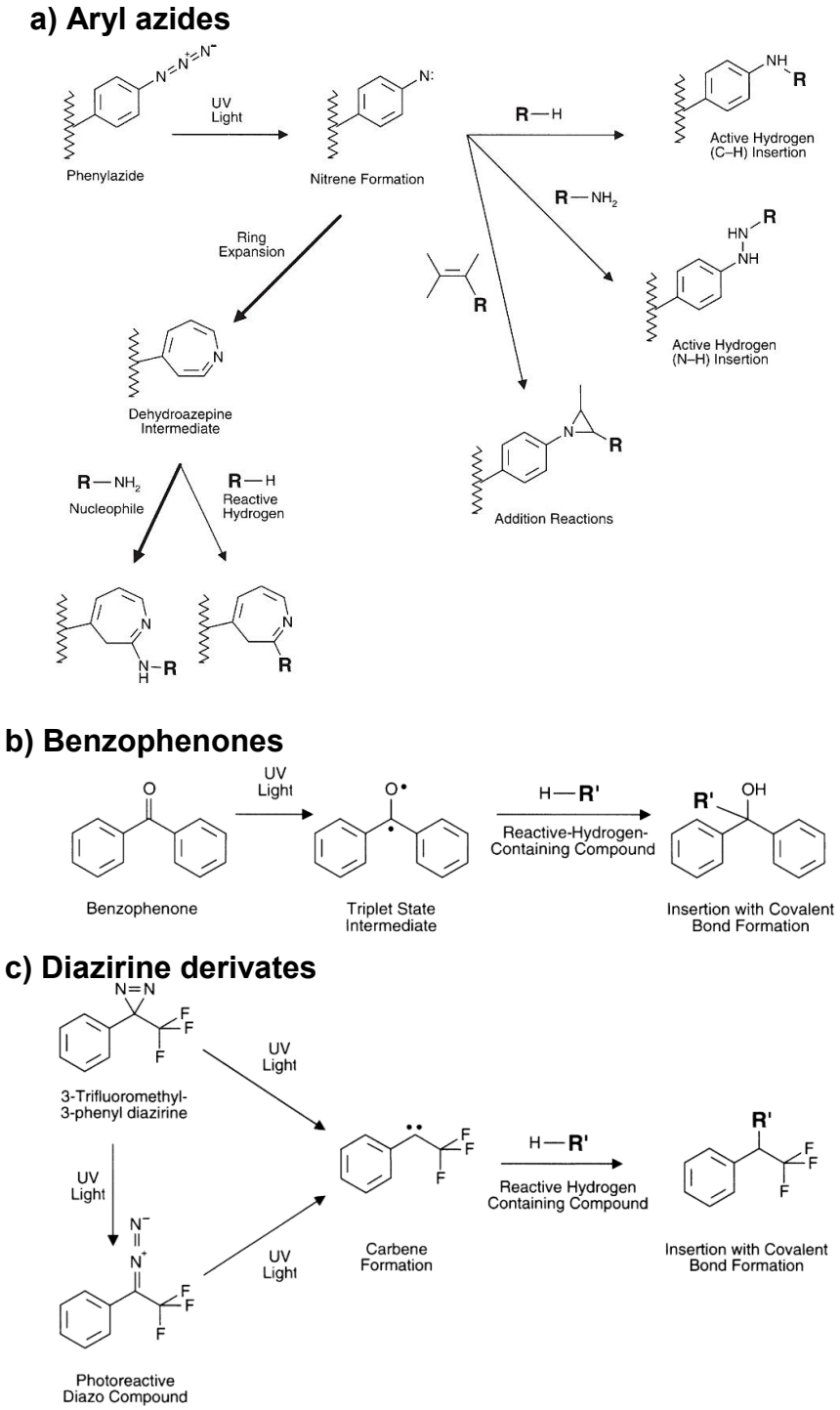


***Figure 26**: Photoactivation of a) aryl azide, b) benzophenone and c) diazirine photolabeling groups. Adapted from*[150].

The previous chapter provides a brief overview of the conjugation techniques commonly used for protein labelling purpose. However, for interested readers in bioconjugation is highly recommended to focus on publications like Hermanson[150] and Begley[164], where the exhaustive description of bioconjugation techniques for other biomolecules is covered in detail.

1.2.2.3 Reactive radicals

The reactive radicals are the third group of covalent labelling reagents, beneficially combining features of HDX and chemical amino acid-selective reagents. As in the HDX, radical labelling occurs on microseconds time scale, enabling "freezing" information of structural intermediates[171,172]. Besides HDX, where the exchange occurs across all amino acid residues (except proline), radicals have limited specificity, but its crucial benefit arises from the irreversible nature of reaction products. Compared to chemical labelling, the reaction speed is also profitable because most of the chemical probes reaction time is mostly in minutes or more, in which modified residues can disrupt the native protein structure and create labelling artefacts[173].

Nowadays, **hydroxyl radicals (•OH)** are the most commonly used radicals for footprinting purpose. Hydroxyl radicals are naturally occurring reactive oxygen species (ROS). Their size and hydrophobicity properties are similar to the water, making them ideal for probing proteins SASA[174]. Due to ROS's known involvement in the cells pathogenesis, great interest has been placed to determine rate constants[175] and reaction mechanisms[176] of amino acids reactions with •OH. The found rate constants for all amino acids are sorted in the following **Table 1** by their reactivity.

Table 1: *Rate constants for reactions between free amino acids and •OH sorted by decreasing reactivity*[177].

substrate	rate constant (M^{-1} s^{-1})	pH	substrate	rate constant (M^{-1} s^{-1})	pH
Cys	3.5×10^{10}	7.0	Pro	6.5×10^{8}	6.8
Trp	1.3×10^{10}	6.5–8.5	Gln	5.4×10^{8}	6.0
Tyr	1.3×10^{10}	7.0	Thr	5.1×10^{8}	6.6
Met	8.5×10^{9}	6–7	Lys	3.5×10^{8}	6.6
Phe	6.9×10^{9}	7–8	Ser	3.2×10^{8}	~6
His	4.8×10^{9}	7.5	Glu	2.3×10^{8}	6.5
Arg	3.5×10^{9}	6.5–7.5	Ala	7.7×10^{7}	5.8
cistine	2.1×10^{9}	6.5	Asp	7.5×10^{7}	6.9
Ile	1.8×10^{9}	6.6	Asn	4.9×10^{7}	6.6
Leu	1.7×10^{9}	~6	Gly	1.7×10^{7}	5.9

The hydroxyl radicals can be generated in many ways. However, only methods used for biomolecules studies will be closely described in the upcoming paragraph.

a) Fenton reaction

The first method for the generation of hydroxyl radicals dates back to 1894 when Fenton[178] published his work focused on the reactivity of H_2O_2 catalyzed by Fe(II). Although the method is named after the Fenton, its mechanism was described in detail by Haber *et al.* in 1932[179,180] (yes, that Fritz Haber who invented the synthesis of ammonia). In general, the Fenton reaction can be summarized by following the reaction scheme[181].

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{\bullet}OH + OH^-$$

$$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + HO_2^{\bullet} + H^+$$

$$H_2O_2 + {}^{\bullet}OH \rightarrow HO_2^{\bullet} + H_2O$$

$$H_2O_2 + {}^{\bullet}OH \rightarrow HO_2^{-\bullet} + H_2O$$

$$Fe^{3+} + HO_2^{\bullet} \rightarrow Fe^{2+} + O_2 + H^+$$

$$Fe^{3+} + O_2^{-\bullet} \rightarrow Fe^{2+} + O_2$$

$$Fe^{2+} + {}^{\bullet}OH + H^+ \rightarrow Fe^{3+} + H_2O$$

In general, the •OH radicals are generated from Fe(II)-EDTA mixed with H_2O_2. Additionally, ascorbate or thiols are added to the mixture securing a reverse reduction of Fe(III) to Fe(II). As an example of a labelling experiment with Fenton-reaction, Sharp *et al.*[182] published work focused on mapping SASA of apomyoglobin.

Although the Fenton reaction is accessible to the broad scientific community (minimal instrumentation requirements) and easy to perform, its use is rarely seen in current publications due to its reaction rate. To achieve sufficient •OH labelling yield, the Fenton catalyzed reaction must be kept at least for several minutes[183].

b) Synchrotron water radiolysis

Synchrotron's capability of water radiolysis for the generation of •OH leading to oxidation of biomolecules was proven in 1997 by Bianca *et al.*[184]. From its name, •OH are generated directly from the water by irradiation of highly energic X-rays. The following equation is summarizing the generation of •OH.

$$H_2O \xrightarrow{h\nu} H_2O^+ + e^-_{dry} \xrightarrow{H_2O} H_3O^+ + \cdot OH + e^-_{aq}$$

Unlike the Fenton reaction, the labelling experiments using synchrotron are conducted on a microsecond timescale. However, the main limitation of this technique comes from the limited availability of synchrotron accelerators.

c) Photolysis of hydrogen peroxide

Last but not least, the photolytic generation of hydroxyl radicals is based on the homophotolytic decomposition of H_2O_2 described by the following reactions[185].

$$H_2O_2 \rightarrow 2^{\bullet}OH$$

$$^{\bullet}OH + H_2O_2 \rightarrow H_2O + HO_2^{\bullet}$$

$$HO_2^{\bullet} + H_2O_2 \rightarrow H_2O + O_2 + {}^{\bullet}OH$$

In 2004 Sharp *et al.*[186] published the first work proving the H_2O_2 photolysis approach for the labelling purpose of proteins. In this publication, lysozyme in PBS buffer containing 15% H_2O_2 was irradiated by a UV lamp for 60 minutes (**Figure 27**), yielding gradual oxidation through time. After one hour, evidence of protein unfolding was observed via extensive oxidation of originally buried amino acids.

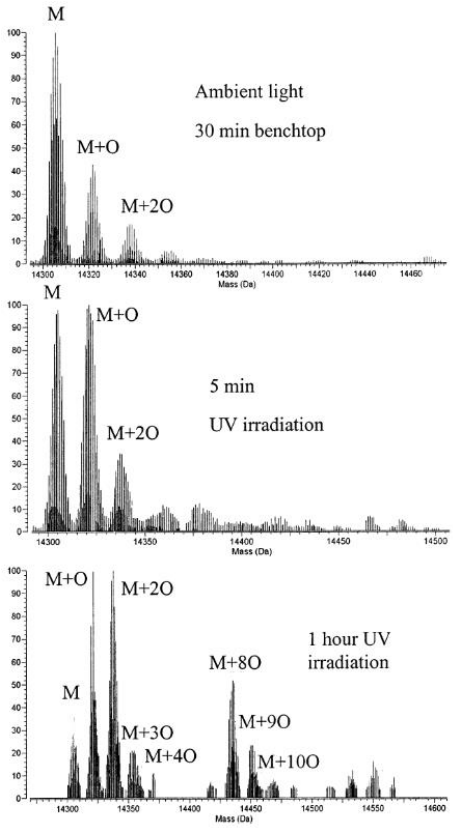


***Figure 27**: Spectra of oxidized lysozyme. The upper spectrum represents lysozyme subjected to H_2O_2 without UV irradiation (control). The middle spectrum reveals oxidation produced by •OH after 5 minutes of UV photolysis. The bottom spectrum shows structure collapse due to the overoxidation (M+10xO) after the 1-hour UV irradiation. Adopted from*[186].

This pioneering publication led other groups to the use of laser serving as a high-energy light source. In 2005 simultaneously, the applications of Nd:YAG (266 nm, fourth harmonic)

and KrF excimer (248 nm) lasers were shown, proving their ability to generate •OH from low concentrated H_2O_2 (0.3% and 0.04%) within a short labelling pulse of 3-5 ns and 17 ns, respectively[171,187]. However, the second publication using the KrF laser introduced a revolutionary experimental setup utilizing a glass syringe loaded with a sample premixed with H_2O_2 (**Figure 28**). Additionally, the introduction of radical scavenger (readily oxidizable amino acids, e.g. histidine, glutamine and methionine) to the H_2O_2 and sample mixture was described, affecting oxidation rate[179]. The syringe was then connected to quartz fused silica tubing with the section of removed coating, creating a small transparent window for the laser beam. Subsequently, this setup was placed in an external syringe pump, and the flow rate was set accordingly to the repetition of the laser to secure that the flowing sample undergoes only one laser pulse[188]. Afterwards, the oxidized sample was collected in a collection tube to which catalase was added, decomposing unreacted H_2O_2, which could potentially cause artificial oxidation during subsequent sample handling[172].

The described experimental setup using quench-flow microfluidics became a milestone in fast radical footprinting, which was then termed **Fast Photochemical Oxidation of Proteins (FPOP)**[171].

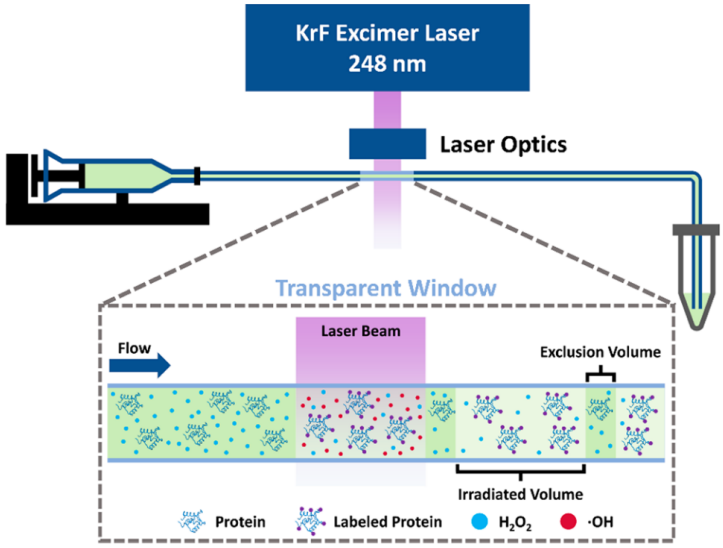


Figure 28: *Scheme of capillary flow apparatus adapted for the FPOP experiment. Adapted from*[181].

The previously described capillary setup allowed photooxidation experiments of various proteins. However, FPOP experiments of proteins prone to oxidation by H_2O_2 were not achievable due to induced oxidation during sample preparation (mixing sample and H_2O_2 in syringe) prior to laser footprinting. Therefore, the implementation of mixing micro-Tee allowed to mix protein with H_2O_2 just before (~ tens of ms) laser irradiation[189].

After the FPOP experiment, the collected sample is usually processed by a standard bottom-up approach. The utilization of top-down MS has not been published yet; however, proof of concept is presented on ubiquitin and myoglobin in publication Yassaghi *et al.* attached to this thesis (under peer-review).

The FPOP employs highly reactive hydroxyl radicals, leading to the oxidation products of 19 of 20 amino acids (except glycine). The reactivity of each amino acid was studied in detail by Xu *et al.*[177], revealing reactivity order: Cys > Met > Trp > Tyr > Phe > cystine > His > Leu, Ile > Arg, Lys, Val > Ser, Thr, Pro > Gln, Glu > Asp, Asn > Ala > Gly. In addition, the oxidation mechanisms for each amino acid have been described separately, revealing multiple oxidation pathways leading to heterogeneous oxidation products. Reaction heterogeneity is illustrated in **Table 2**, where oxidations products are described with their mass shifts observable within MS. As could be seen from this table, the most common oxidation results in the formation of one or more hydroxy- groups (+16 Da, +32 Da, etc.)[177].

Table 2: *Primary oxidation products and the corresponding mass changes for amino acids side chains*[177].

Side chain	Modification and mass change
Cys	sulfonic acid (+48), sulfinic acid (+32), hydroxy- (to serine, −16)
Met	sulfoxide (+16), sulfone (+32), aldehyde (−32) Trp
Trp	hydroxy- (+16, + 32, + 48, etc.), pyrrol ring-open (+32)
Tyr	hydroxy- (+16, + 32, etc.) ing
Phe	hydroxy- (+16, + 32, etc.)
cystine	sulfonic acid (+48+H), sulfinic acid (+32+H) His
His	oxo-(+16), ring-open (−22, −10, +5)
Leu	hydroxy- (+16), carbonyl (+14)
Ile	hydroxy- (+16), carbonyl (+14)
Arg	deguanidination (−43), hydroxy- (+16), carbonyl (+14)
Lys	hydroxy- (+16), carbonyl (+14)
Val	hydroxy- (+16), carbonyl (+14)
Ser	hydroxy- (+16), carbonyl (−2, or +16-H2O)
Thr	hydroxy- (+16), carbonyl (−2, or +16-H2O)
Pro	hydroxy- (+16), carbonyl (+14)
Gln	hydroxy- (+16), carbonyl (+14)
Glu	decarboxylation (−30), hydroxy- (+16), carbonyl (+14)
Asp	decarboxylation (−30), hydroxy- (+16)
Asn	hydroxy- (+16)
Ala	hydroxy- (+16)
Gly	N/A

Conclusively, the FPOP beneficially uses •OH radicals as probes having water-like size and hydrophobicity. Achieving high labelling efficiency of nearly all amino acids within a few microseconds makes the FPOP an ideal footprinting technique for studying fast dynamic events (protein folding/unfolding, protein complex formation/dissociation, etc.).

On the other hand, the minor disadvantage is attributed to the predominant reaction product yielding in a mass shift of +16 Da, which can be wrongly misinterpreted as a modification if the protein was already oxidized prior to the labelling reaction. Hence, control experiments devoid of labelling are unavoidable, giving quantitative evidence of background oxidation. This issue leads scientists to the development of new radical probes providing different mass shifts than those obtained by FPOP. The most relevant radical probes will be discussed in the following chapter[106,181].

Fluoroalkyl radicals

Although fluorine is the 13th most abundant element, its occurrence in the biologically relevant compounds is rare (only five compounds with the C-F bond were reported)[190]. Hence, the insertion of covalently bound fluorine to biomolecules makes it an ideal atom for labelling purpose.

Trifluoromethyl radical alkylation was firstly introduced in 1949 by Emeléus and Haszeldine[191]. However, the generation of radicals from CF_3I was challenging, requiring reactions to be performed in the gas phase and under high temperatures. Much later, Umemoto's[192] and Togni's[193] reagents were developed (**Figure 29a, c**), requiring activation by transition metals or Lewis acids[194–196]. In 1991 Langlois *et al.*[197] introduced a new stable trifluoromethylation reagent trifluoromethanesulfinate (CF_3SO_2Na) (**Figure 29b**) which utilized t-butylhydroperoxide (TBHP) as a radical inducer, providing sufficient trifluoralkylation yield under mild reaction conditions[198,199].

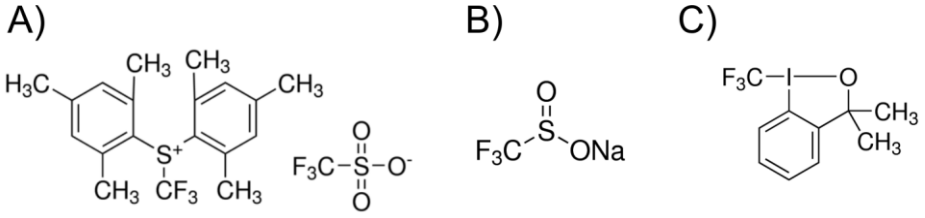


***Figure 29**: Representative compounds used for radical trifluoralkylation. a) Umemoto's reagent, b) Langlois reagent, c) Togni's reagent*

In 2017 Cheng *et al.*[200] described an innovative approach for generating $\bullet CF_3$ from Langlois reagent through KrF laser photolytic decomposition of hydrogen peroxide (as in the FPOP) and formation of primary •OH radicals. Secondary $\bullet CF_3$ radicals were then formed through the following radical propagation shown in **Figure 30**.

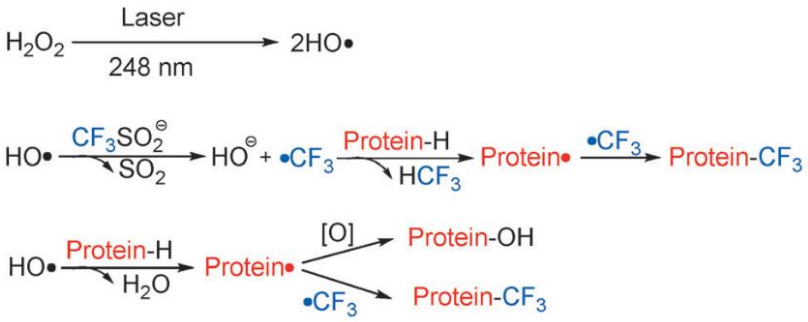


Figure 30: *Proposed mechanism of trifluoromethylation protein footprinting. Adapted from*[200].

For such an experiment, the quench-flow apparatus previously built for FPOP was utilized. Prior to the experiment, protein, hydrogen peroxide, and Langlois reagent were mixed and filled into the Hamilton syringe connected to silica tubing with a transparent window for laser beam (KrF 248 nm). As the reaction product, Cheng *et al.*[200] observed mass shifts of +67.987 Da corresponding to •CF_3 for 18 out of 20 amino acids (excluding Met and Cys). However, preferential reactivity toward aromatic acids was shown (Trp, Tyr, His, and Phe). One of the first applications of laser-initiated radical trifluoromethylation was shown on the holo/apomyoglobin, studying structural change in the presence/absence of the heme group. The footprinting experiment revealed peptide 80-96 containing His81, His82, and His93 to be significantly more modified in the apomyoglobin than holoform (**Figure 31**). This observation is in excellent agreement with previously published NMR[201] and top-down HDX data[202], where apomyoglobin's helix F (structure in **Figure 31a**) is not structured in the absence of the heme prosthetic group.

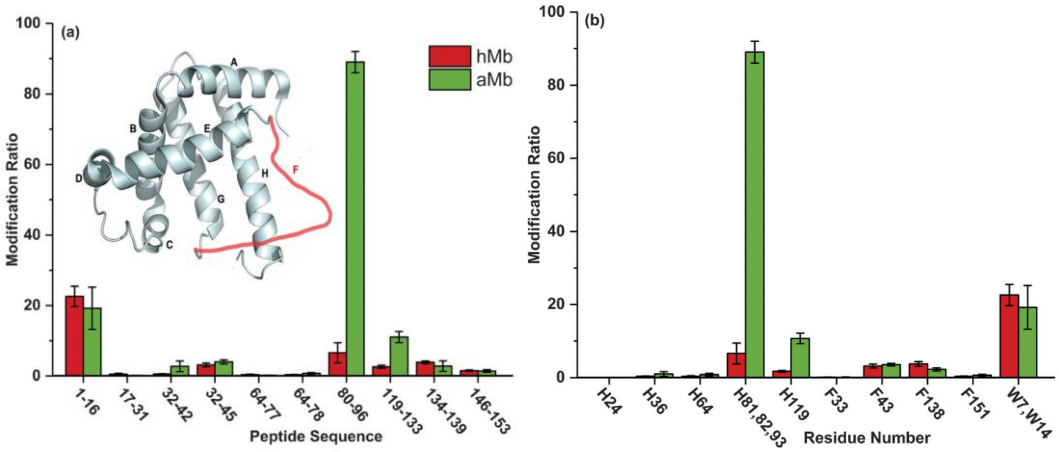


Figure 31: *Comparison of modification extents for holomyoglobin (red) and apomyoglobin (green) on a) peptides and b) single residues levels. The most significant modification extent difference was observed for 80-96 peptide, corresponding to the heme-binding F-helix location in holomyoglobin. In the absence of heme, the helix is in unstructured form (show in the picture). Adapted from*[200].

The previous paragraph revealed the application of Langlois (CF_3SO_2Na) reagent for protein footprinting purpose. However, in 2008 Caponea e*t al.*[203] used hypervalent iodine reagents (Togni's reagents) to modify pharmacologically significant Octreotide (octapeptide) selling under commercial name *Sandostatin*®. In this publication, various Togni's reagents were tested, and •CF_3 radicals were generated by various methods. However, none of them was applicable for the labelling of native proteins due to the presence of organic solvents or high temperatures[196].

Nevertheless, in 2019 Rahimidashaghoul *et al.*[204] published a new type of Togni's reagents, but most importantly, a novel method for the efficient generation of •CF_3 under biocompatible conditions. The proposed reaction mechanism introduces sodium ascorbate as a critical component that serves as a radical inducer (**Figure 32**).

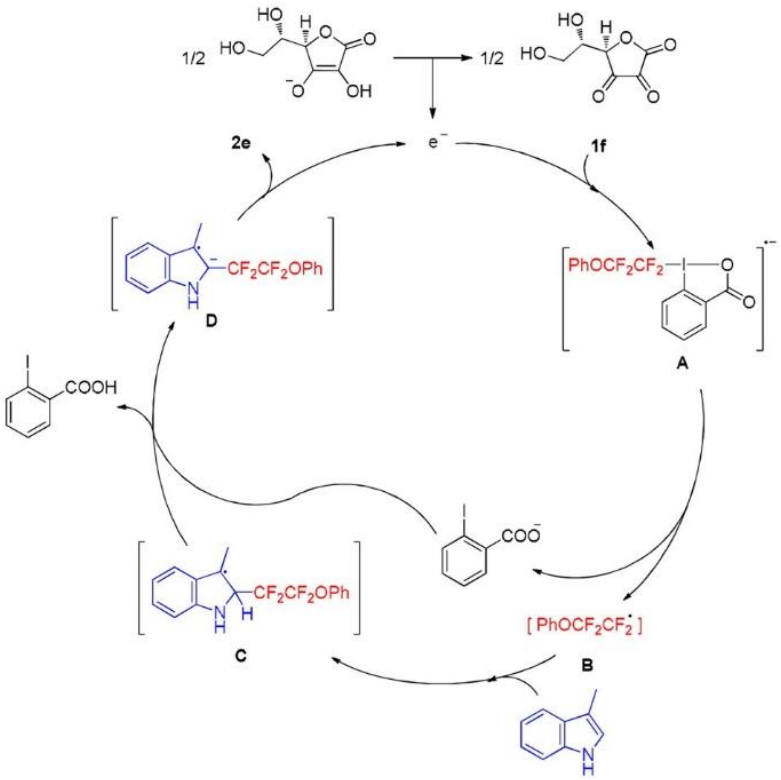


Figure 32: *Proposed reaction mechanism utilizing sodium ascorbate as a radical inducer for fluoralkylation of 3-methylindole with Togni's reagent. Adapted from*[204].

The reaction of •CF_3 with individual amino acids revealed the following reactivity order: Trp >> cysteine > Tyr > Phe > His. Surprisingly, tryptophan yield in one mono(trifluoromethylated) and the other two isomeric bis(trifluoromethylated) products. Additionally, two mono(trifluoromethylated) products of phenylalanine were identified. Finally, the practical application was demonstrated on carbonic anhydrase I and ubiquitin,

where modified aromatic amino acids were identified after a reaction time of 1 hour and 15 minutes, respectively[204].

Aside from described footprinting techniques utilizing hydroxy radicals and trifluoromethyl radical many others exist (**Figure 33**). Namely, sulfate radical anions (**SO_4^{-}•**)[205]; carbonate radical anions (**CO_3^{-}•**)[206]; carbenes (**R-C:-R**)[207]; iodine radicals (**I•**)[208], generated by multiple approaches (UV light, thermolysis, transition metal catalysis etc.), are providing researchers almost unlimited toolbox of footprinting techniques. On top of that, unique reactivity (very specific **I•** or non-specific **OH•**) complementing each other makes radical footprinting an ideal tool for each researcher interested in the biologically exciting cases (structure, interactions dynamics, immobilizations etc.)[181].

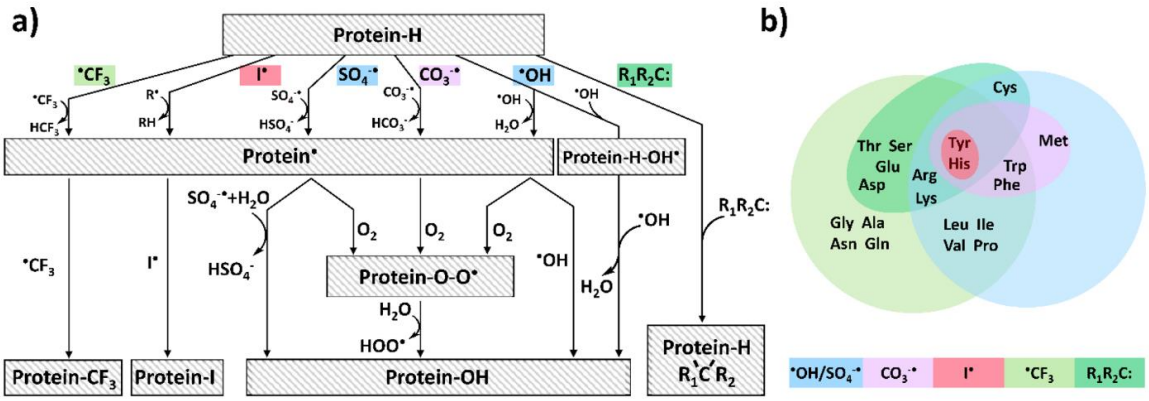


Figure 33: *Summarizing illustration of selected footprinting radicals a) presenting simplified reactions pathways and b) radicals selectivity. Adapted from*[181].

1.2.3 Chemical cross-linking

Last but not least, chemical cross-linking completes a broad repertoire of structural MS techniques. Its first use is dating back to 1958 where insulin was subjected to difluorodinitrobenzene forming intramolecular products, which reveals insulin polypeptide subchains composition[209]. However, utilization of cross-linking with mass spectrometry (XL-MS, CXMS, or CLMS) was not published until 1991, when Farmer and Caprioli[210] published the article where glutaraldehyde cross-linker was used for cross-linking of protein multimeric complexes. The first use of chemical cross-linking for structural characterization (interatomic distance constraints) was published in 2000 by Young *et al.*[211] In that publication, the advantageous use of lysine-lysine XL-MS was utilized to determine the tertiary structure of bovine fibroblast growth factor (FGF)-2. Since then, XL-MS became an emerging field utilizing new and already known chemistry to develop new cross-linkers and cross-linking strategies.

In general, a chemical probe creates covalent bonds between two reactive groups in its proximity. For this purpose, cross-linkers (also known as chemical rulers) are used, composed of various reactive groups with unique specificity, connected by a spacer of different lengths, providing different distance restraints[212].

In principle, cross-linkers could be divided into four basic groups by their composition (**Figure 34**). The most commonly used are homobifunctional cross-linkers whose use two identical reactive groups connected with the variable linker. The next type of cross-linkers, known as heterobifunctional, has a linker and two various reactive groups. The third group, named zero-length, are cross-linkers that directly connect two molecules without adding any atom to form a bond. Lastly, the trifunctional cross-linkers are known, utilizing the added reactive group for various usages such as cross-linked peptides enrichment, the introduction of an affinity tag, etc.[213].

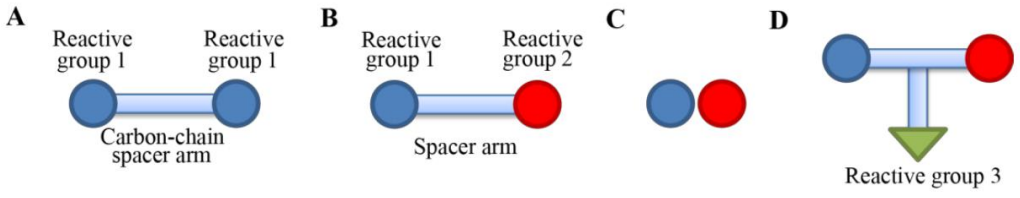


Figure 34: *Different cross-linker types: a) Homobifunctional cross-linker, b) heterobifunctional cross-linker, c) zero-lenght cross-linker d) trifunctional cross-linker. Adapted from*[214].

Cross-linker's selectivity is driven by reactive groups that follow the same reactivity pathways as reagents used for chemical labelling purposes. However, chemical labelling results in only two defined products (unlabelled protein; protein + probe(s)), while chemical cross-linking yields three different reaction products, significantly increasing mixture complexity (**Figure 35**). The structural information is encoded in Type 1 (intrapetide) and Type 2 (interpeptide) cross-links. However, Type 0 (dead-end) formed through reaction with only a single cross-linker group still carrying relevant information about solvent accessibility similar to those obtained by chemical labelling[215,216].

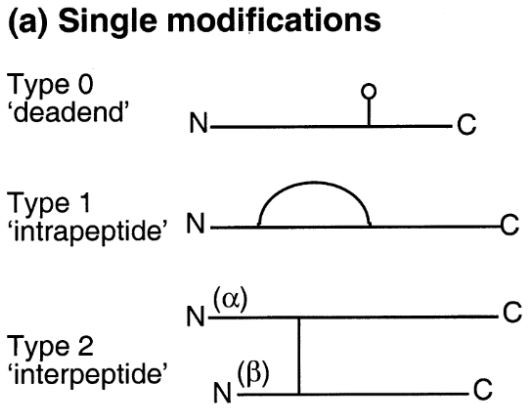


Figure 35: *Nomenclature of proteolytic peptides (α, β) and cross-linking products. Adapted from*[215].

As was shown, Type 0 cross-links are in their nature labels created in a way described in chapter 1.2.2.2 chemical labelling. In the same chapter, **Figure 21** presents the percentage of the solvent-exposed amino acid residues revealing their accessibility toward chemical labels. The most surface-exposed amino acid was found to be lysine, making it especially susceptible to labelling reagents such as N-hydroxysuccimide (NHS) esters, which is the ideal reactive group for lysine-lysine cross-linkers[150].

In the mid-1970s, the first experiments introducing NHS-based cross-linkers were published, revealing their practical use for structural studies[217,218]. A simplified cross-linking scheme on NHS-based cross-linker is illustrated in **Figure 36a**. However, NHS-based cross-linker's poor water solubility led Stratos[219] to synthesize new NHS cross-linker analogues whose succinimido ring was modified by -SO_3H group. The introduction of the sulfo-NHS group gives cross-linkers a charge and enough polarity to make them water-soluble. Additionally, sulfo-NHS were found to be more resistant to unwanted hydrolysis leading to cross-linker inactivation (**Figure 36b**)[220].

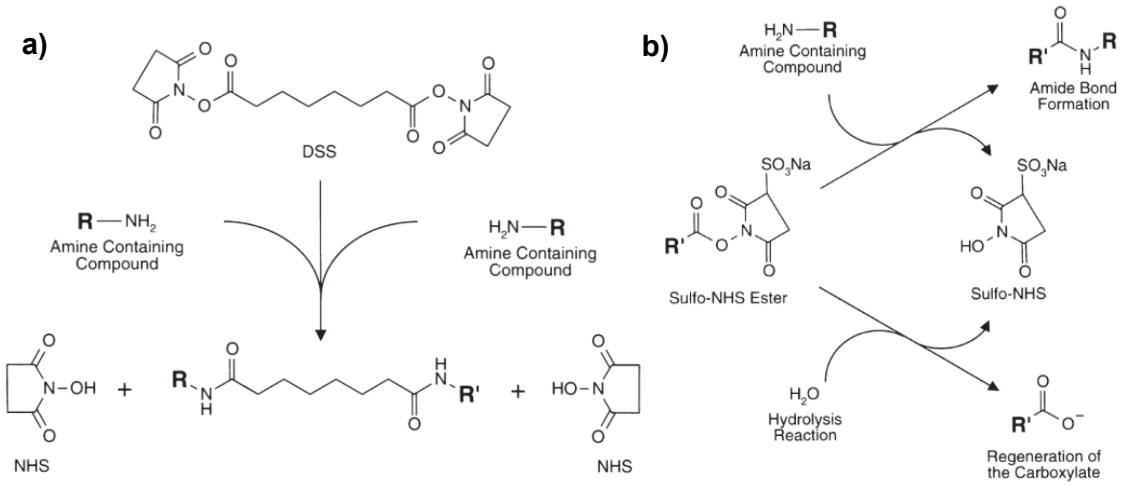
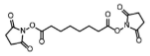
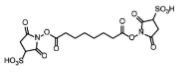
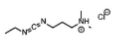
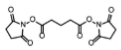
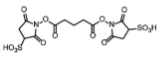


***Figure 36**: a) scheme of chemical cross-linking of primary amines (lys-lys) by disuccinimidyl suberate (DSS). b) Reactions of sulfo-NHS with primary amine and water. Adapted from*[150].

Since then, NHS-based cross-linkers with various spacers became the gold standard for lysine-lysine cross-linking (also reactivity towards tyrosine, serine, and threonine was reported[221,222]). Through years of development, many types of cross-linkers also became commercially available. A brief overview of commonly used cross-linkers with the additional features is presented in **Figure 37**.

non-cleavable crosslinker	structure	features
DSS		isotope-coded available commercial spacer-length 12 Å
BS3		isotope-coded available commercial spacer-length 12 Å
EDC		zerolength commercial
DMTMM		zerolength commercial
DSG		isotope-coded available commercial spacer-length 7.7 Å
BS2G		isotope-coded available commercial spacer-length 7.7 Å

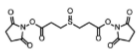
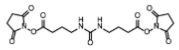
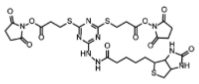
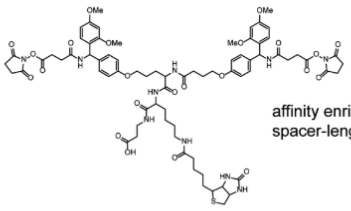
cleavable crosslinker	structure	features
DSSO		commercial spacer-length 10.1 Å
DSBU		commercial spacer-length 12.5 Å
CBDPS		commercial affinity enrichable isotope coded spacer-length 15 Å
PIR		affinity enrichable spacer-length 43 Å

***Figure 37**: Commonly used non-cleavable and MS/MS cleavable cross-linkers. Adapted from*[223].

Although the previous text was focused more on cross-linker reactive groups, spacer parameters are also playing a crucial role that should be considered before conducting any experiment. The most basic parameter is linker length, typically determined by the number of carbons connecting two reactive groups. The linker length influences spatial resolution and data density (the longer the spacer is, the more residues can be linked). Nevertheless, longer spacers reduce restraint values for structural modelling because they set only upper-bound distance restraints[224].

Additionally, the spacer can be isotopically coded by stable isotopes (e.g., ^{2}H, ^{13}C, ^{15}N, ^{18}O)[225,226]. A cross-linked sample by an equimolar mixture of light/heavy cross-linker yields cross-linked peptides with mass shifts corresponding to the mass difference between light/heavy cross-linker (e.g. disuccinimidyl glutarate; DSG H_4/D_4; Δm = 4.0313 Da). This effect is easily observable in mass spectra by the presence of specific mass shifts ("doublets"), which simplifies cross-linked peptides identification (**Figure 38b**) and allows the use of a quantitative cross-linking (QXL-MS) approach[227,228]. However, retention differences within reverse-phase liquid chromatography (RPLC) were observed, especially for deuterated cross-links, causing problematic quantification[229]. This problem can be solved by using ^{13}C or using MS/MS cleavable cross-linkers described in the following text.

The third aspect of the spacer is the presence of labile moiety (sulfoxide[230], urea moieties[231]), which undergoes cleavage within the gas phase fragmentation induced by collisions (CID) or electrons (ETD/ECD) (**Figure 38c**)[223]. The MS^2 fragmentation is yielding in the formation of a unique pattern of a "doublet of 26 u difference doublets" (Δm = 26 Da for urea-based cross-linkers), giving an evidence of type 2 cross-links. Such, approach based on reporter doublets simplifies cross-links identification and improves false-positive assignments[223].

Most recently, the use of isotopically labelled MS/MS cleavable cross-linkers was published by Ihling *et al.*[232]. This combined approach is improving cross-links identification for highly complex mixtures such as complete proteome (e.g. *Drosophila* embryo extract).

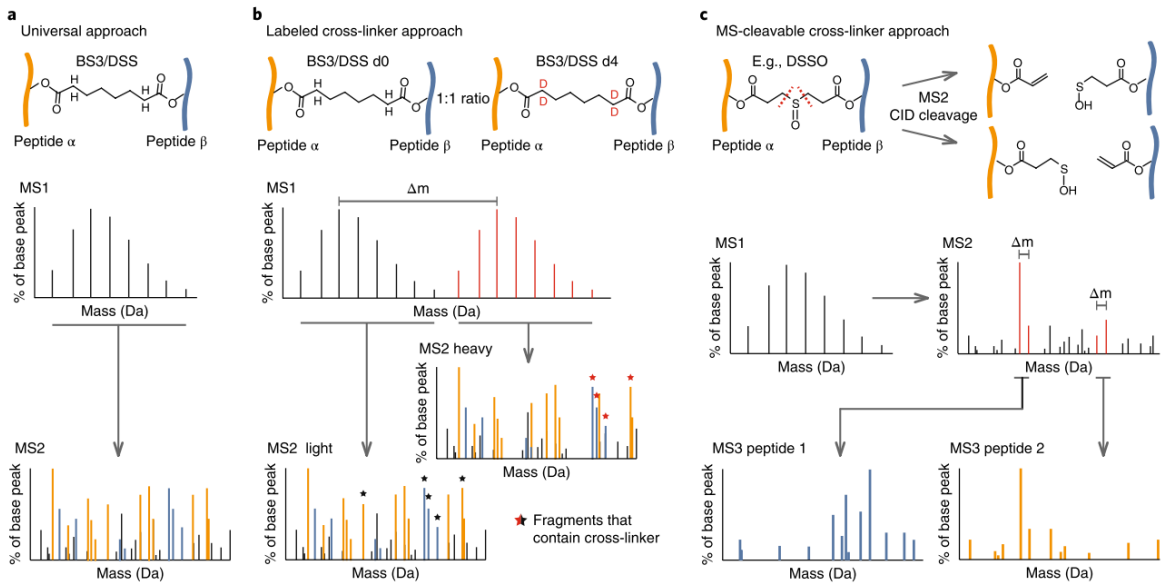


Figure 38: *CL-MS approaches for a) standard, b) isotopically coded, and c) MS/MS cleavable cross-linkers. Adapted from*[224].

Aside from the most common homobifunctional lysine-lysine cross-linkers, many others exist, utilizing combined chemistry principles described in the previous chapter 1.2.2.2 chemical labelling. For example, homobifuctional cross-linkers targeting carboxyl groups through EDC activation and secondary reaction with various dihydrazide spacers are used[233,234].

On top of that, heterobifunctional cross-linkers allow endless combinations of reactive groups, yielding unique cross-linkers specificity towards different amino acids or even different biomolecules like DNA/RNA[235]. For illustration, two heterobifunctional cross-linkers succinimidyl 4,4'-azipentanoate (SDA) and succinimidyl-[4-(psoralen-8-yloxy)]-butyrate) (SPB) are shown in **Figure 38**. As can be seen, both cross-linkers possess NHS group selective towards lysines. The second group in SDA is diazirine that can be photoactivated by UV irradiation (330-370 nm), forming nonspecifically reacting carbenes[236]. Further, the SPBs reactive group is 8-methoxypsoralen, capable of intercalation into double-stranded DNA. Upon UV irradiation (>350 nm), psoralen undergoes cycloaddition with thymines, forming a particular DNA-protein cross-linking product[237,238].

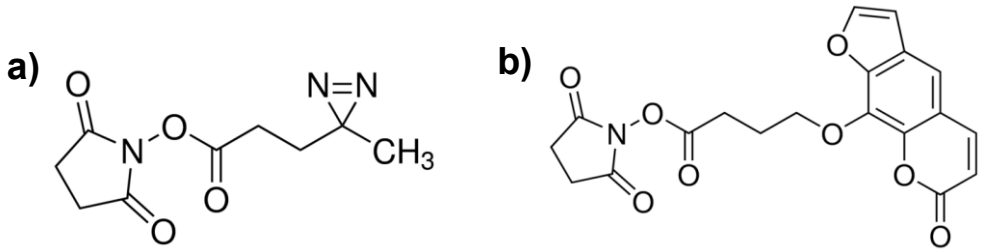


Figure 38: *Heterobifunctional UV activated cross-linkers a) SDA and b) SPB*

As was shown, chemical cross-linking has became an essential technique within structural MS. The diversity of commercially available cross-linkers with almost unlimited selectivity makes it a powerful tool in the hand of scientists interested in the HOS of various biological systems. Despite the cross-linkers diversity and applicability of the standard proteomics bottom-up workflow, its main bottleneck currently lies in the field of data processing and interpretation. Especially, proteome-wide cross-linking datasets are still challenging even though isotopically coded or MS/MS cleavable cross-linkers with additional enrichment step are used. Hence, current improvements are focused primarily on developing algorithms capable of automatic cross-links assignments[239].

2. Aims of the study

The thesis aim was to extend the versatility of covalent labelling and chemical cross-linking techniques for the characterization of proteins and protein-protein complexes. The primary aim was focused on techniques where the use of microflow reactors could achieve short labelling times.

The specific goals were to:

- Develop quench-flow apparatus for Fast photooxidation of proteins by the excimer laser. Subsequently, utilized top-down and bottom-up approach to determine oxidation sites of proteins and their complexes.
- Adapt quench-flow apparatus for labelling of proteins and protein complexes by Togni's reagents.
- Test influence of cross-linker leaving group polarity toward its selectivity (NHS vs sulfo-NHS group).
- Develop a cross-linking approach for the determination of assembled intermediate filaments.

3. Methods

The attached manuscripts include a detailed description of experimental methods, chemicals and instruments, sufficient for the full reproduction of all performed experiments. Thus, this chapter includes only the list of the most commonly used methods.

List of used research methods:

- Chemical cross-linking
- SDS-PAGE for confirmation of cross-linked proteins
- Fast photooxidation of proteins and radical fluoralkylation in quench-flow setup
- MS analysis of protein primary structure and their post-translational modifications (FT-ICR MS, timsTOF Pro direct infusion ESI and nanoESI or in LC-MS and LC-MS/MS mode)

4. Results and Discussion

The thesis aim was to increase the potential of chemical labelling and chemical cross-linking techniques for the characterization of proteins and protein-protein complexes. Therefore, the result section is divided into parts dedicated to pulse labelling techniques and chemical cross-linking.

4.1 Development of pulse labelling techniques

To prepare a short labelling experiment, samples can be prepared simply by pipetting. However, at a certain point (in the order of seconds), the limit of the shortest feasible time that can be reproducibly prepared is reached (limited sample mixing, slow transfer of reacting liquid components, imprecise pipetting, etc.). At this point, the possibility of converting the whole reaction into a microfluidic flow system, which can provide reaction times in the order of milliseconds to several seconds with high reproducibility, is becoming very advantageous.

To build a flow reactor for labelling purpose, a few essential items are needed, namely: precise syringe pumps, syringes, capillaries, micromixers and basic knowledge of liquid flow.

The flow rate as the basic parameter can be set on every pump. However, for the subsequent calculations, it is suitable to work with linear velocities, which are obtained by conversion according to the equation:

$$v = \frac{4F}{\pi d^2}$$

Where v is the linear velocity (mm s^{-1}), F is the flow rate (µl min^{-1}), and d is the tubing diameter (mm).

This parameter is almost sufficient for the basic design of quench-flow apparatus for the FPOP experiment used in the following publications. However, an additional parameter, the width of the transparent window (L in millimetres), is needed to calculate and secure that the flowing sample will be irradiated by a single laser shot securing the protein sample from radiation damage. For such purpose, the following equation can be used yielding in the repetition rate of laser (f in Hz), which secures that sample undergoes only a single irradiation pulse.

$$f = \frac{v}{L}$$

However, the calculated repetition rate is valid only if the plug-like flow is present. Nevertheless, the most common type of flow in capillaries is the laminar flow with different velocities at the capillary centre and near walls (Reynolds numbers <<2000). This type of flow has a parabolic flow profile leading to a liquid delay near the walls[146].

In FPOP experiments, the setting of exactly calculated frequency leads to the multiple exposures of part of the sample. For this reason, lower laser frequencies are used to achieve at least 20% exclusion volume between each laser pulses (**Figure 39**)[240].

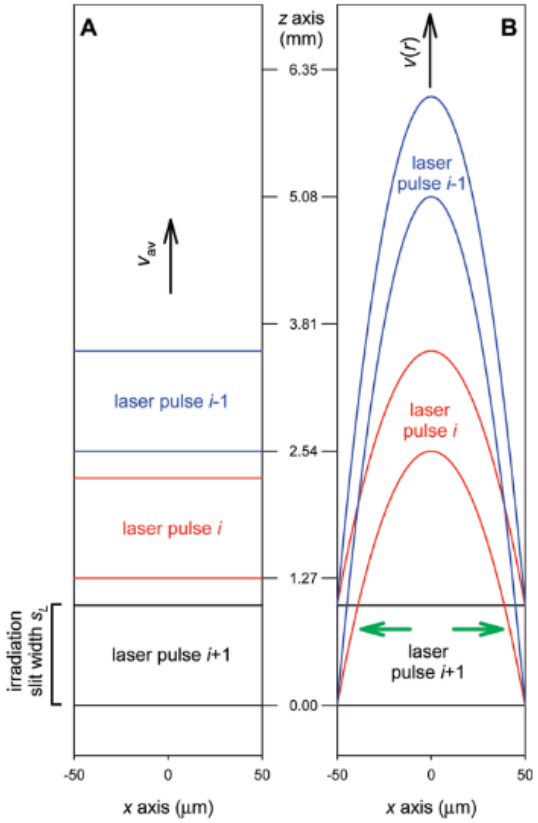


***Figure 39**. Schematic cross-sections of a flow tube illustrate the temporal evolution of three consecutive volume elements exposed to laser-induced oxidative labelling: a) plug flow; b) laminar flow. Adapted from*[240].

Based on the described parameters, a quench-flow reactor for the FPOP footprinting experiments conducted in **Publication 1, 2 and 3** was designed and built. The scheme and experimental setups are presented in **Figure 40**.

The experimental quench-flow setup was composed of a dual-channel pump for syringes filled with sample and H_2O_2 and a single-channel pump for a syringe filled with the quencher (methionine). The glass syringes were connected by micromixing T-pieces, ensuring mixing in the required ratios determined by both the flow rates ratio (F_1/F_2) and the fused silica

capillary diameters (d_1/d_2). The removal of the capillary coating created the transparent window for the excimer laser beam.

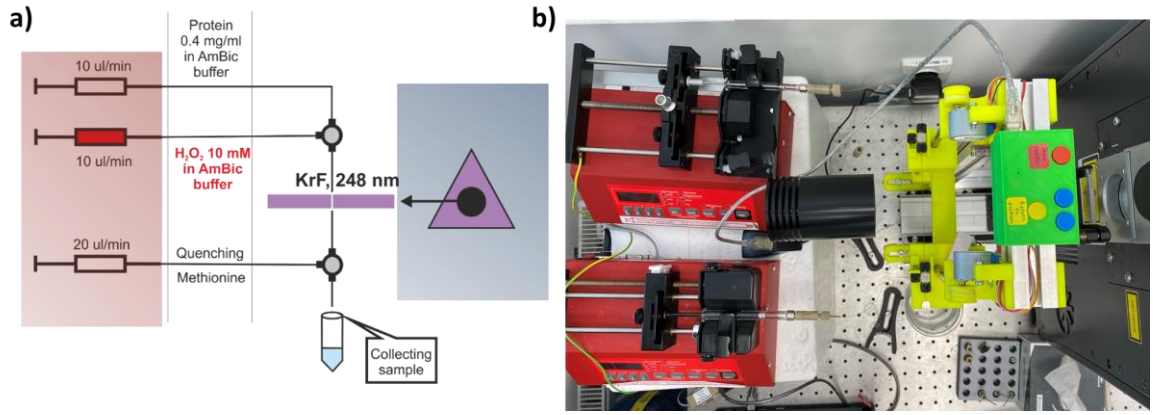


***Figure 40**: a) Scheme of quench-flow apparatus for FPOP footprinting. b) Experimental setup.*

Based on the FPOP pulse labelling setup, I've decided to adapt it for experiments utilizing Togni reagents for a novel technique termed Fast Fluoro Alkylation of Proteins (FFAP). In the FFAP, Togni reagents serve as the source of $\bullet CF_3$ radicals generated immediately upon mixing with ascorbic acid, which serves as a radical inducer. Furthermore, the duration of the radical reaction can be controlled by the subsequent addition of a quencher (Tryptophan).

For FFAP experiments, a different approach had to be used to achieve a short reaction time than in FPOP, where the radicals are generated by a very short laser pulse (20 ns). The labelling time within FFAP experiments was controlled by the time which sample (mixture of protein, Togni reagent and ascorbic acid) spent in the reaction capillary (delay line) before mixing with the quencher. The following equation sums up all parameters that need to be known for the calculation of the reaction labelling time.

$$t_{labelling} = \frac{V_d}{F_1 + F_2} = \frac{\pi r_d^2 L}{F_1 + F_2}$$

In this equation, V_d represents delay line volume, F_1 and F_2 flow rates of sample+Togni reagent and ascorbic acid, respectively. r^2_d represents the radius of the delay line capillary, and L represents its length.

Based on this equation, I've designed and built a quench-flow reactor that provides labelling times of less than 1 second. A detailed scheme of quench-flow apparatus used in **Publication 4** for FFAP labelling experiments providing 3-second labelling pulses is presented in **Figure 41.**

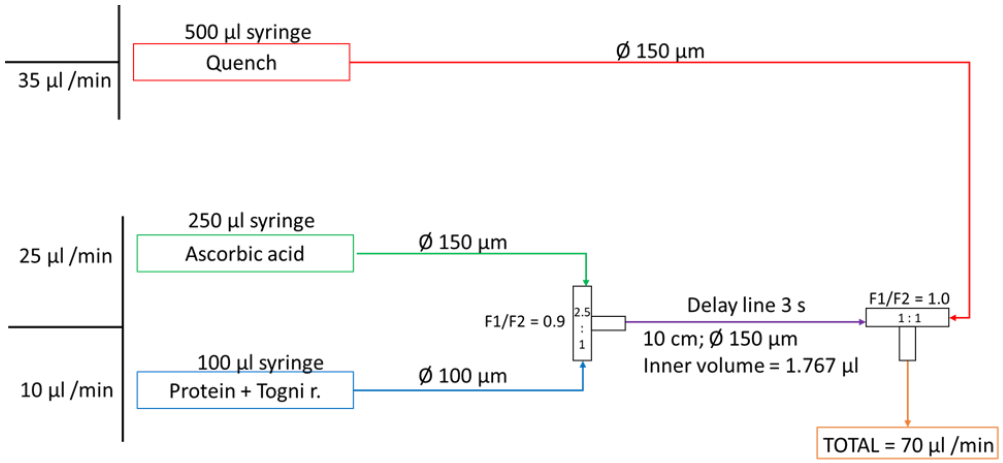


***Figure 41**: Scheme of pulse labelling apparatus used for FFAP experiments.*

4.1.1 Publication 1

Title: Fast Photochemical Oxidation of Proteins and Top-Down Mass Spectrometry: The Effect of Collision-Induced Dissociation, Electron-Transfer Dissociation and Electron-Capture Dissociation on Product-Ion Spectra. (*under review*)

Structural mass spectrometry is evolving field introducing novel techniques like Fast photochemical oxidation of proteins. In most cases, oxidized proteins are analyzed by the standard bottom-up approach. However, this approach derives information about total oxidation from peptides that yield from various protein populations with different modifications. Therefore, this publication proposes a top-down approach, where the oxidation sides are determined only from singly-oxidized protein population that reflect the initial native protein conformation.

First, FPOP experiments were conducted with well-studied model protein ubiquitin. The ubiquitin was oxidized in the quench-flow apparatus. **Figure 42** shows the extent of protein oxidation. Subsequently, the most intense singly-oxidized protein population (charge state 10+) was isolated by quadrupole and subjected to CID, ETD and ECD fragmentations.

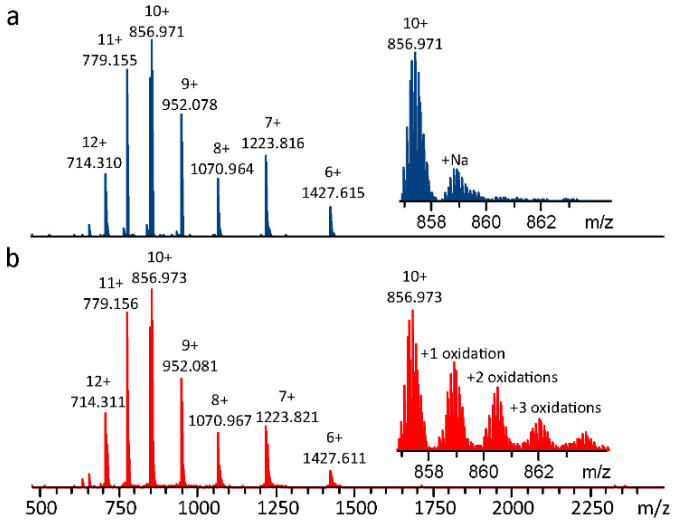


***Figure 42**: Full mass spectra of the non-oxidized (a) and oxidized (b) ubiquitin. The inset shows the mass spectra corresponding to the 10+ charge state of respective proteins. Comparison of both spectra illustrates the presence of singly, doubly, and triply oxidized ubiquitin in the sample subjected to the FPOP.*

The fragmentation spectra were used to calculate the oxidation extents of each fragment, pinpointing oxidized amino acids residues. Based on the fragmentation efficiency, ETD yielded the lowest amount of fragment ions that resulted in insufficient spatial resolution of the experiment. On the contrary, the CID fragmentation achieved better spatial resolution, however significant neutral loss of water complicated the data evaluation process. The best spatial resolution was accomplished by EDC, where nearly a single residue resolution was reached (**Figure 43**).

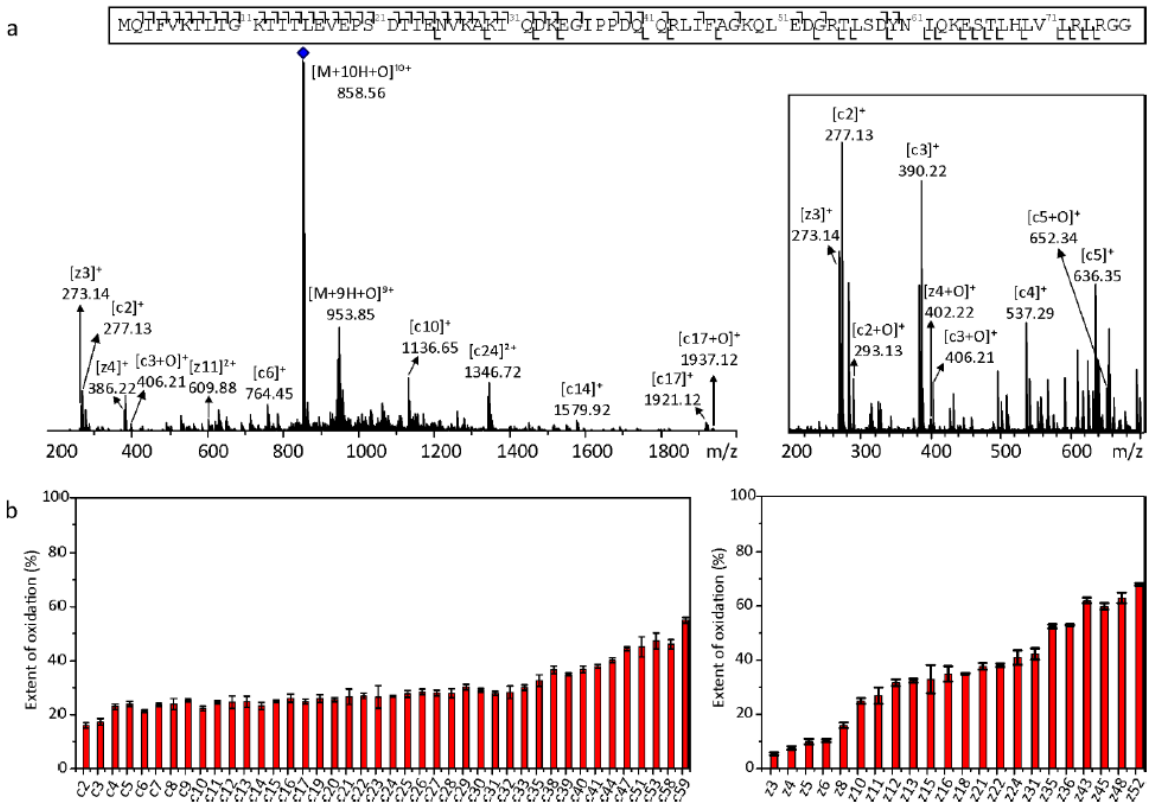


***Figure 43**: ECD spectrum of FPOP labelled ubiquitin. a) Overview of the whole spectrum obtained upon isolation of 10+ charge state singly oxidized ubiquitin (m/z 858.56) indicated with blue diamond, with zoom-in ECD spectrum of the m/z range 200-650. b) Extent of FPOP oxidation for ubiquitin on the fragment level. Red bars indicate oxidized fragment ions. Error bars are showing the standard deviations of three independent MS measurements.*

Subsequently, CID, ETD and ECD fragmentations of singly-oxidized ubiquitin were performed for additional charge states (9+, 11+ and 12+). Surprisingly, the CID fragmentation of different charge states revealed the dependence of the oxidation extent on the charge state.

Multiple continuous accumulation of selected ions (multi CASI) was used to overcome this phenomenon, allowing simultaneous isolation of several charge states within one experiment. Multi CASI/CID and multi CASI/EDC experiments of singly-oxidized ubiquitin (9+, 10+, 11+ and 12+ charge states) were tested. Unexpectedly, multi CASI/CID fragmentation resulted in fewer fragments than its CID analogue due to the significant water loss of oxidized residues. In contrast, multi CASI/EDC enhanced sequence coverage and yielded more stable oxidation extents.

In the next step, the top-down approach using multi CASI/CID and multi CASI/EDC was applied to monitor conformational changes. As a model protein, horse hearth holomyoglobin

and its apoform (without heme group) were subjected to FPOP. Afterwards, singly-oxidized charge states (15+, 16+, 17+, 18+, and 19+) of holomyoglobin or apomyoglobin were simultaneously fragmented, resulting in a satisfactory amount of fragment ions (combination of multi CASI/CID and multi CASI/EDC fragment ions). Subsequently, oxidation extents were calculated for each fragment ion of holo/apomyoglobin separately. If the difference in the extent of oxidation (holo vs apoform) exceeded the specified threshold value, this change was considered significant and highlighted in myoglobin's structural model (**Figure 44**). As a result, we obtained a myoglobin model where red residues were more oxidized in apoform and green more oxidized in holoform. More oxidized residues in apoform (red) are preferentially in the heme-binding cavity giving evidence of structural changes upon release of the heme group. Additionally, our top-down data are in excellent agreement with previously published bottom-up FPOP experiments conducted by Vahidi *et al.*[241].

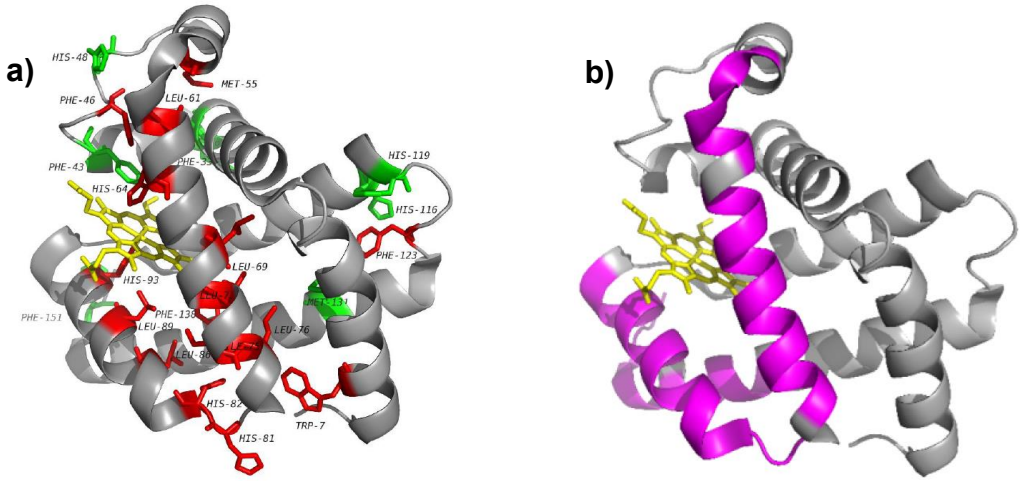


***Figure 44**: a) Myoglobin structure (PDB: 1WLA)*[242] *with oxidized residues labelled by FPOP and detected by a top-down approach. The residues featuring significantly increased oxidation levels are highlighted for apomyoglobin (red) and holomyoglobin (green). b) Myoglobin structure by classical bottom-up FPOP approach*[241]*. Differences in oxidation for apoform are highlighted in violet.*

In this publication, we demonstrated the novel utilization of a top-down approach for analyzing FPOP experiments. Unfortunately, ETD fragmentation yielded poor fragmentation efficiency and was not helpful for further experiments. The CID provided sufficient fragmentation; however, water loss occurred preferentially on oxidized fragment ions, biasing the obtained data. The EDC fragmentation provided the best spatial resolution (nearly single amino acid resolution), making it the ideal fragmentation technique available

for top-down FPOP experiments, especially in combination with multi CASI. Finally, we demonstrated a novel top-down approach to determine structural changes of holo/apoform of myoglobin.

4.1.2 Publication 2

Title: Benefits of Ion Mobility Separation and Parallel Accumulation−Serial Fragmentation Technology on timsTOF Pro for the Needs of Fast Photochemical Oxidation of Protein Analysis

In the **Publication I**, FPOP top-down approach was successfully presented for small proteins (ubiquitin and myoglobin). **Publication 2** and **publication 3** are dedicated to test the benefits of timsTOF Pro (quadrupole time-of-flight (QTOF) coupled with trapped ion mobility) for standard FPOP experiment. Unlike standard QTOFs, ion mobility separation before MS/MS adds another dimension of separation, which allows additional separation of precursors with the same *m*/*z* but the different shape of molecule (like peptide with one oxidation (+16 Da) on two different residues). Additionally, this novel instrument is equipped with "parallel accumulation−serial fragmentation" (PASEF) technology, which synchronizes quadrupole and tims elution time to achieve fast sequencing speed (>100 Hz) and nearly 100% duty cycle[243]. All these features can be beneficial for FPOP experiments where almost all amino acids (except glycine) can be oxidized, with various oxidation products (described in chapter 1.2.2.3 **Table 2**).

In this publication, Haptoglobin 1-1 composed of Hpα and Hpβ subunits, was subjected to oxidation by hydroxyl radicals generated from H_2O_2 by KrF laser (248 nm). Also, a control experiment was performed by mixing Hp with H_2O_2 without laser irradiation (detection of background oxidation caused by H_2O_2). One hundred sequence coverage was achieved by subsequent MS analysis. Additionally, samples subjected to radical footprinting revealed significantly more modifications compared to samples without laser irradiation. Specifically, 6 and 55 modifications were found on 5 and 15 types of amino acids of Hpα and Hpβ, respectively (**Figure 45a**).

Subsequently, modifications were plotted on the structural model of Hp 1-1 (based on PDB 4WJG), revealing the location of oxidized amino acids near the protein surface (**Figure 45b**). This observation nicely fitted the calculated solvent-accessible solvent area concerning amino acids reactivity, where exposed and reactive amino acids

(aromatic residues and Met) yield to a greater extent of modification than those buried into protein structure.

a)

residue	position	
	Hpα	Hpβ
Cys	111	266, 340
Trp		175, 284, 398
Tyr	114, 125	272, 280, 386
Met		263, 300
Phe		173, 205, 290, 292, 339
His		166, 208, 295
Arg		261
Ile		233
Leu	127	167, 206, 221, 230, 296
Val	124	231, 275, 278
Pro		219, 268, 302
Thr		293, 338
Ser		282
Lys		170, 178, 228, 235, 270, 291
Asp	122	397

b)

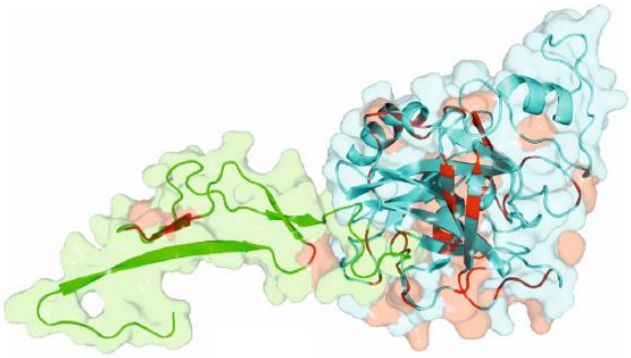


***Figure 45**: a) Amino acids modified with hydroxyl radicals during the FPOP experiment b) Structural model of Hp 1-1 (based on PDB 4WJG) with highlighted modified residues identified in the laser-irradiated sample. The α subunit is in green color, the β subunit is in cyan, and the modified residues are red. Respective light colors represent the solvent-accessible area of the protein surface.*

Beneficial separation by trapped ion mobility was demonstrated on selected singly-oxidized peptide DYAEVGR (m/z = 825.3568) in **Figure 46**. In this case, the same peptide had two retention times of 5.6 and 6.5 minutes, yielding identical fragmentation patterns pointing to modification of tyrosine. However, ion mobility revealed different $1/k_0$ values, indicating the modification of different tyrosine aromatic ring positions. A similar effect was observed for various peptides. Thus, thanks to ion mobility separation, sub-residue resolution of oxidized peptides was achieved.

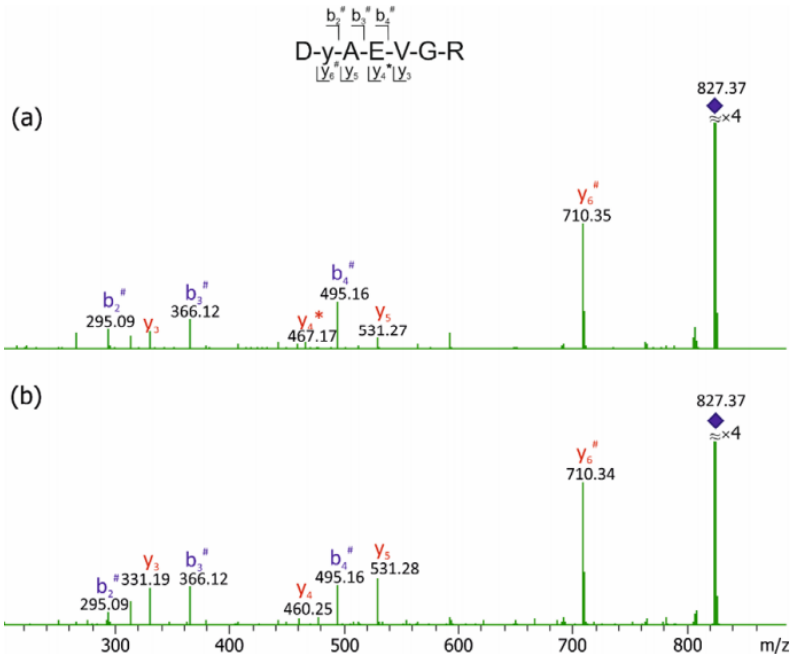


Figure 46: *Fragmentation spectra of the peptide 271-277 (m/z 825.3568) with a retention time of 5.6 (a) and 6.5 min (b). A precursor ion is labelled with blue rhombus. # - oxidized fragment ions, * - loss of ammonia. The modified residue is labelled with a lowercase letter.*

Demonstration of PASEF technology was shown on peptide KQLVEIEK (*m/z* 501.78), where five different residues were found to be oxidized (**Table 3**). Multiple oxidation sites were supportively indicated by distinct peptides retention times and also different ion mobilities.

The key benefit of PASEF technology is that the fragmentation spectra for peptides can be obtained over a wide dynamic range of intensities. In the case of mentioned oxidized peptide, the lowest and the highest intensities differ by more than 3 orders of magnitude. Despite that, fragmentation spectra were sufficient to identify the modification site precisely, even for low abundant oxidation products.

Table 3: *List of detected modifications in the KQLVEIEK peptide*

−10lgP	z	RT	$1/k_0$ range	intensity	oxidation site	AScore
38.14	2	2.97	0.8442–0.8628	147,992	L3	26.31
41.03	2	3.79	0.8345–0.8531	1224	V4	30.46
36.83	2	5.83	0.8286–0.8472	16,835	K8	26.31
43.59	2	6.12	0.8487–0.8673	25,697	K1	36.05
38.07	2	7.15	0.8717–0.8903	5250	I6	32.28
53.33	2	6.31	0.8583–0.8769	1,479,460	NA	NA

The presented publication has shown the suitability of a novel timsTOF Pro instrument for the FPOP experiments. Due to the complexity of the oxidized peptide products, ion mobility separation was beneficially used, adding another dimension of separation. Thanks to IMS, the sub-residue resolution of oxidations was achieved (apparently *ortho*, *meta* position of tyrosine the aromatic ring). Additionally, PASEF technology allowed the precise location of oxidized residues for peptides varying in concentration, covering almost four orders of magnitude dynamic range.

4.1.3 Publication 3

Title: Hydroxyl radical footprinting analysis of a human haptoglobin-hemoglobin complex. (*accepted*)

A subsequent publication is a follow up of the previously published article describing the combination of the FPOP approach and the timsTOF Pro instrument. Thanks to the instrument's features, a single-residue resolution of oxidized amino acids of the human Hp-Hb complex was achieved, uncovering the interaction interphase of complex (tetramer composed of Hpα, Hpβ and Hbα, Hbβ subunits).

Although the function of this protein complex has been thoroughly investigated, its high-resolution structure has not yet been satisfactorily resolved. Up to date, there are only two X-ray structural models of this complex in the PDB database, while one represents structural model (2.9 Å) of porcine analogue[244], and the second is the human Hp-Hb complex, but only as a multimeric complex with a *Trypanosoma brucei brucei* receptor[245].

The Hp, Hb and their complex were separately subjected to FPOP. For a bottom-up data interpretation, a comprehensive list of all possible outcomes of protein hydroxyl radical footprinting was created using already published data considering the reactivity of each amino acid (**Table 4**).

Thanks to the ion mobility separation and PASEF technology, PEAKS X+ software assigned a sufficient number of modified residue distributed over the Hp and Hb sequences resulting in excellent spatial resolution of the experiment (**Table 5**).

Table 4: *List of FPOP modifications defined in PEAKS X+ software.*

Modification	Targets
FPOP oxidation / + 15.994915	M, W, Y, F, H, L, I, V, P, R, K
FPOP dioxydation / + 31.989828	C, M, W, Y, F
FPOP trioxydation / + 47.984745	C, W
FPOP carbonyl / + 13.979265	L, I, V, P, R, E, Q, H
FPOP Met-32Da / - 32.008457	M
FPOP His+5Da / 4.9789	H
FPOP His-10Da / -10.031968	H
FPOP His->Asp / -22.03197	H
FPOP His->Asn / -23.01598	H
FPOP Arg_deguanidination / - 43.053432	R
FPOP decarboxylation / - 30.010565	Q, N
FPOP ST_carbonyl / - 2.01565	S, T

Table 5: *Number of identified peptides by PEAKS X+ software. All – all identified peptides; Mod. – number of modified peptides with a localization score ≥ 20; Cert. – number of modified peptides with at least one unambiguously modified residue.*

	Hbα			Hbβ			Hpα			Hpβ		
	All	Mod.	Cert.	All	Mod.	Cert.	All	Mod.	Cert.	All	Mod.	Cert.
Hb	149	78	34	218	142	50	-	-	-	-	-	-
Hp	-	-	-	-	-	-	89	29	2	282	116	58
Hb-Hp	294	144	49	335	184	52	169	59	2	429	184	61

The quantitative extend of modification for single protein and proteins in complex encoded the Hp-Hb interacting regions. For illustration, **Figure 47** represent such a result for Hpβ where the orange columns correspond to oxidation found for "free" Hpβ and blue columns correspond to oxidation when the Hpβ was oxidized in complex

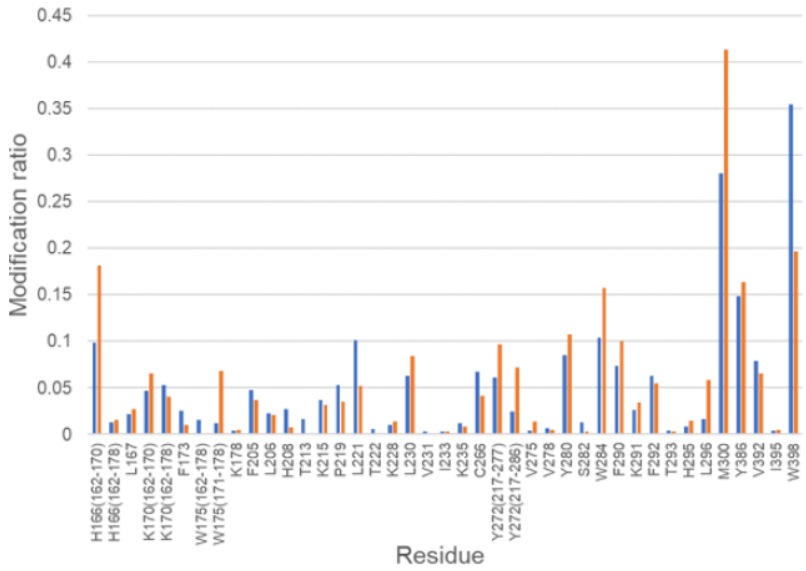


Figure 47: *Modification ratios at the residue level (orange columns – "free" protein, Blue columns – Hb–Hp complex). Residues with multiple modifications, like His, are presented as the sum of all detected modifications contributing to the overall oxidation extent.*

As can be seen, **Figure 47** revealed differences within oxidation level for "free" Hpβ and Hpβ bound in the complex. For example, amino acids H166, W284 and M300 represent more oxidised residues in "free" form, giving evidence of protection upon binding to complex assembly. On the contrary, residues L221 and W398 shown the opposite effect (deprotection). For graphical evaluation, oxidation differences ("free" vs complex) for all complex subunits were plotted on structure (based on PDB: 4WJG), revealing position within the Hp-Hb complex (**Figure 48**). Residue coloured in red and blue represented less and more oxidized amino acids (relative to the complex), respectively. As can be seen, red residues are located at the interaction interphase (e.g. Hpβ: H166, W284, M300; Hbβ: W37), explaining their lower oxidation extent upon complex formation (protection).

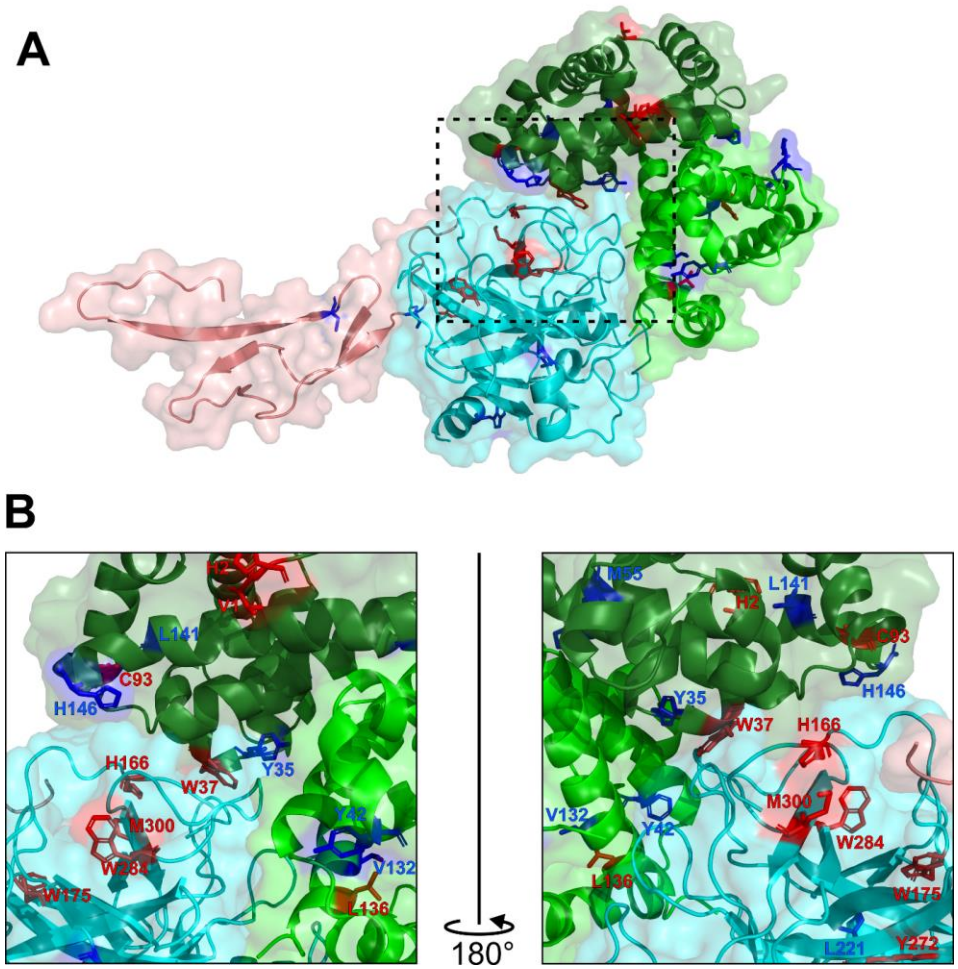


Figure 48: *Structural model of Hb–Hp complex with highlighted residues oxidized by the FPOP experiment. a) Overview of a tetrameric complex composed of α,βHp and α,βHb subunits. b) Zoom of interaction interface. Red residues were less oxidized (protected) upon the formation of complex, showing evidence of the interaction interface. Blue residues represent amino acids that were more oxidized after the complex's formation. Hpα is in pink, Hpβ is in cyan, Hbα is in green, and Hbβ is in dark green.*

Based on the performed FPOP experiment, three Hp (βH166, βW284, βM300) and three Hb residues (αLeu136, βTrp37, βCys93) were found to be less oxidized (protected) upon formation of the Hp-Hb complex. This observation revealed the interaction interphase of the protein complex. Additionally, our data are in agreement with theoretical calculations performed by Nantasenamat *et al.*[246], where βTrp37 Hb was pinpointed as a "hot-spot" significantly responsible for the formation of the Hp-Hb complex.

4.1.4 Publication 4

Title: Fast Fluoroalkylation of Proteins Uncovers Structure and Dynamics of Biological Macromolecules.

This publication describes the development of a novel footprinting technique termed Fast Fluoro Alkylation of Proteins (FFAP). As labelling probes, fluoroalkyl radicals were generated from hypervalent iodine-fluoroalkyl reagents known as Togni's reagents.

In this publication, specifically acid (**Figure 49a, c**) and alcohol (**Figure 49b, d)** type Togni's reagents were used. Additionally, various types of radicals were generated (•CF_3 vs •CF_2CF_2Im). Due to FFAP radical nature, this method utilized capillary quench-flow apparatus similar to FPOP, providing labelling pulses in the timescale of seconds (**Figure 49e**).

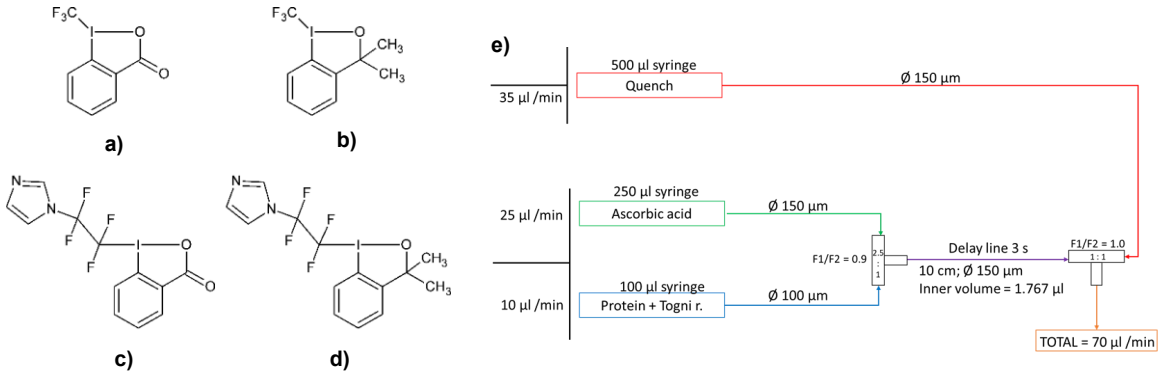


Figure 49: *Structure of cyclic hypervalent iodine-fluoroalkyl reagents used in this study. a) acid type Togni-CF_3 reagent b) alcohol type Togni-CF_3 reagent c) acid type Togni-CF_2F_2Im reagent and d) alcohol type Togni-CF_2CF_2Im reagent. e) Quench-flow apparatus utilized for the FFAP experiment.*

However, unlike the fluoralkylation by Langlois reagents (CF_3SO_2Na) and oxidation by the FPOP, our technique did not require an expensive KrF laser, which makes it feasible for the broad researcher's community.

On top of that, no reaction products were observed for the protein sample and Togni's reagent mixture before the addition of sodium ascorbate (radical inducer), solving the problem of artificial oxidation caused by H_2O_2 within the FPOP experiment.

In this publication, FFAP was utilized on the well-studied system of holo/apomyoglobin and Hp-Hb protein complex.

As the first experiment, the holo and the apo form of myoglobin were separately labelled by all four Togni's reagents shown in **Figure 49**. For the labelling, the quench-flow apparatus was used with calculated parameters of the delay line to yield in a labelling time of three seconds. Subsequently, samples were processed by standard bottom-up protocol, resulting in assigned peptides and peptides with fluoroalkyl modification. Finally, the extents of modification were calculated for each peptide for holo/apoform (red/blue), resulting in a column graph shown in **Figure 50**. In this graph, 19 out of 22 aromatic amino acids were found to be modified after the three-second reaction with acid type Togni-CF_3 reagent.

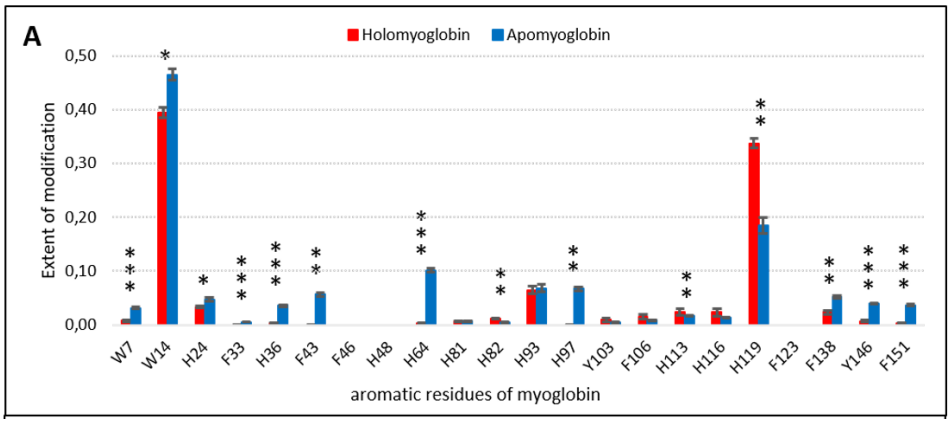


Figure 50: *Quantification of CF_3 modifications of aromatic residues for apo- and holomyoglobin. The radical labelling reaction was performed by acid type Togni-CF3 reagent. Red bars represent the extent of holomyoglobin modification and blue bars represent apomyoglobin modification, P-value (*** - $P < 0,005$; ** - $P < 0,01$; * - $P < 0,05$).*

Differences between apo and holoform were statistically tested by the T-test represented by the asterisk on top of the column. Subsequently, statistically significant differences obtained from all four Togni reagents were plotted on the structure of myoglobin, revealing regions that underwent structural changes (**Figure 51**). For clarity, the regions marked in red were found to be more modified in the apoform, while the blue ones were less modified.

As shown in **Figure 51a,** red regions are in close proximity to the heme group, which explain greater modification extents after the heme group's release. Additionally, similar regions were found to be more oxidized in **Publication 1,** where the same model proteins were used for the FPOP experiment (**Figure 51b**).

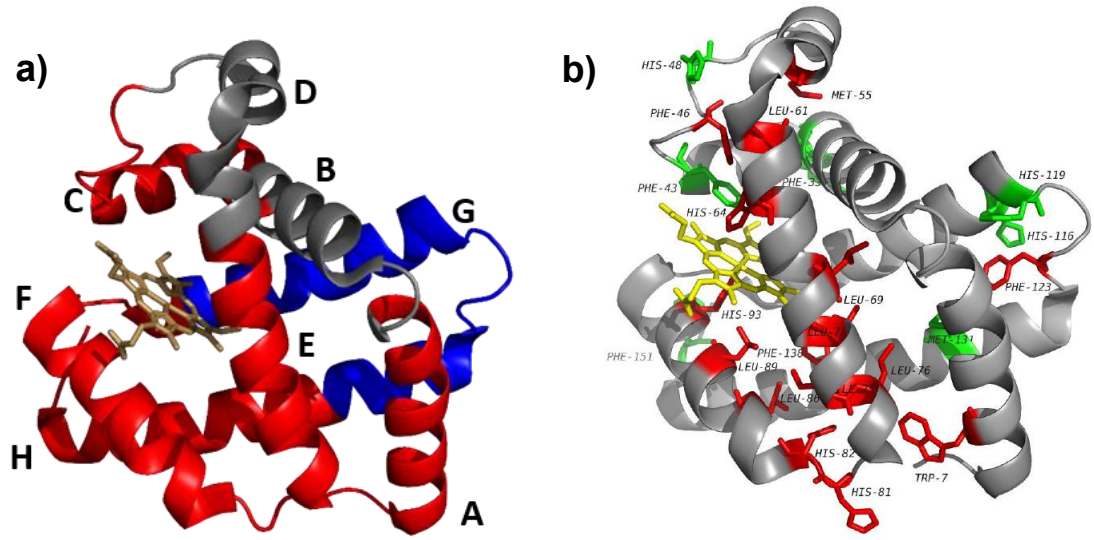


Figure 51: *Visualization of the regions with the highest degree of modification and the difference between apomyoglobin and holomyoglobin obtained by a) FFAP and b) FPOP. The red regions were found to be more modified, while the blue or green regions were less modified in apomyoglobin (absence of heme group).*

For further FFAP experiments, a tetrameric complex of Haptoglobin-Haemoglobin (Hpα, Hpβ and Hbα, Hbβ subunits) was used. Similar to **publication 3**, this experiment aimed to describe the interaction interface between Hp 1-1 and Hbαβ. Hence, sample preparation and data processing for FFAP was almost identical to preparation for the FPOP experiment, so only a brief description is provided.

The Hp, Hb and their Hp-Hb complex were separately subjected to the FFAP with two acid type Togni's reagents, followed by a standard bottom-up approach. The modification extents were calculated for individual subunits while being in the form of "free" proteins (more precisely as Hpαβ dimer and Hbαβ tetramer) and in the tetrameric complex (Hp-Hb). Subsequently, differences were plotted in column graphs (similar to **Figure 50**), and the differences were statistically evaluated using a T-test. Based on the statistics, differences were plotted to the 3D structure of the Hp-Hb complex (**Figure 52**). Blue and red labelled residues were more and less modified upon the formation of the tetrameric complex, respectively. From the zoom in **Figure 52,** red labelled residues were found to be predominantly located between Hpβ and Hbαβ subunits. Due to the location and lower extent of modifications (protection against modifications) of the red-labelled residues, it was possible to determine the probable interaction interface of the Hp-Hb complex. On top of that, the FFAP provided information about interaction interphase, which is in excellent agreement with data presented in **publication 3,** where the FPOP experiment was utilized with the identical Hp-Hb complex.

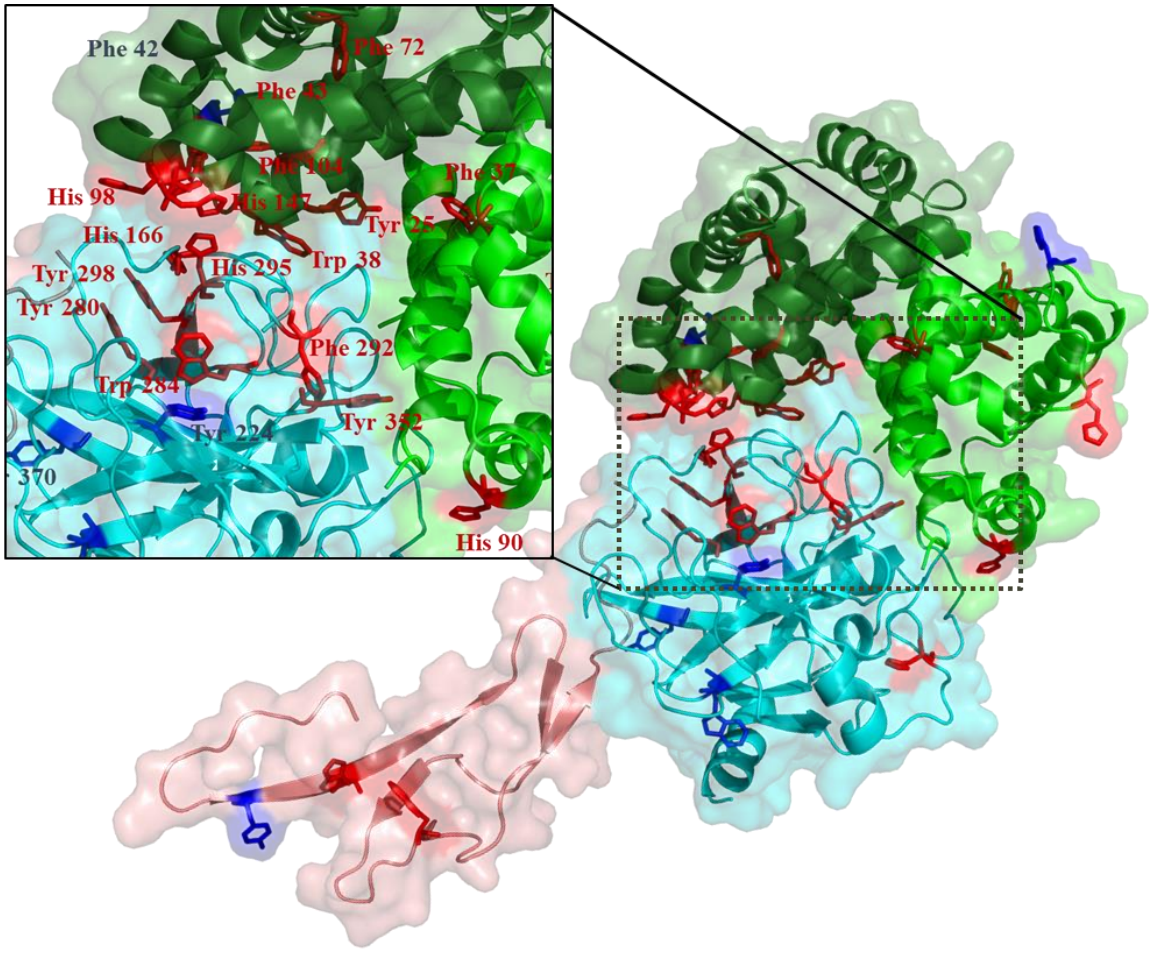


Figure 52: *Structure of Hp-Hb complex with highlighted modified residues after reaction with acid type Togni-CF_2F_2Im reagent. Red residues were less modified, and blue residues were more modified in the Hp-Hb complex. Subunits of the complex are highlighted as follows: Hpα - pink, Hpβ - light blue, Hbα - light green and Hbβ - dark green. Zoom – Interaction interface of Hp-Hb complex.*

In this publication, we presented the novel fluoralkylation strategy of proteins. Due to the similarities with the FPOP, we termed this new technique Fast Fluoroalkylation of Proteins (FFAP). Its feasibility has been demonstrated on model proteins apo/holo myoglobin and the Hp-Hb complex, previously used for structural elucidation by FPOP in **publication 1** and **publication 3**, respectively. As was shown, FFAP provided similar results as FPOP, but without the need for an expensive KrF laser used for the generation of hydroxyl radicals. This fact makes the FFAP an ideal alternative to conventional techniques.

4.2 Chemical cross-linking

4.2.1 Publication 5

Title: Influence of cross-linker polarity on selectivity towards lysine side chains

Nowadays, many cross-linkers with different chemistry mechanisms are used for mapping the structure of proteins or protein complexes. However, often due to their poor solubility in water (hydrocarbon chain spacers, hydrophobic leaving group), non-polar solvents such as DMSO, DMF must be used for their dissolution. To prevent the addition of non-polar solvent to the sample, water-soluble cross-linkers with a polar leaving group such as the sulfosuccinimidyl group (sulfo-NHS) have been synthesized and are commonly available on the market. This publication compares the polarity effect of leaving groups of cross-linking agents disuccinimidyl glutarate (DSG) and bis(sulfosuccinimidyl) glutarate (BS2G).

Based on previous works, sulfo-NHS were preferentially used due to their solubility in the water-based buffers. However, studies proving no difference between NHS and sulfo-NHS cross-linking products were not published. Thus, in this work, we have decided to use DSG and BS2G cross-linkers containing the same linker and their isotopically coded analogues (d6-DSG and d6-BS2G) to reveal the reactive group's polarity effect (**Figure 53**).

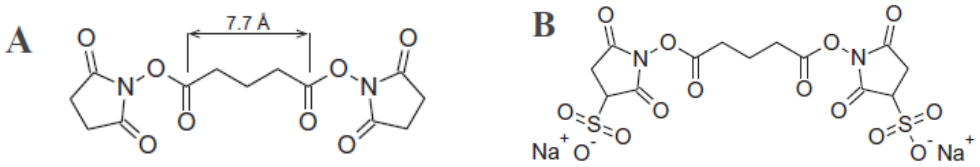


Figure 53: *Chemical structure cross-linkers used in this study. (A) Disuccinimidyl glutarate (DSG) and (B) bis(sulfosuccinimidyl)glutarate (BS2G). BS2G contains two additional sulfonate groups (-SO_3H), which give its negative charge and make it water-soluble.*

In general, cross-linking reaction is performed with cross-linker, and it is isotopically coded analogue in the equimolar ratio (e.g. d0-DSG:d6-DSG). This cross-linking mixture yields cross-linked peptides with a 1:1 ratio of light/heavy linker observed in the mass spectrum as "doublet" with mass shift corresponding to linker isotope difference (in this experiment $\Delta m = 6.0376$ Da). However, to study the effect of leaving groups, equimolar mixtures composed of d0-DSG/d6-BS2G (**sample A**) and d6-DSG/d0-BS2G (**sample B**) were used separately for cross-linking of bovine serum albumin (BSA).

As control experiments, equimolar cross-linkers mixtures with the same leaving group were also used (d0-DSG/d4-DSG and d0-BS2G/d4-BS2G). After that, cross-linked BSA was digested by trypsin and measured via LC-MS/MS.

Surprisingly, cross-linking reactions of BSA by d0-DSG/d6-BS2G and d6-DSG/d0-BS2G yield cross-linked peptides with light/heavy ratios far from the expected 1:1 ratio. Additionally, ratios were reciprocal (**Figure 54**). On top of that, control experiments showed the ideal 50:50 ratios.

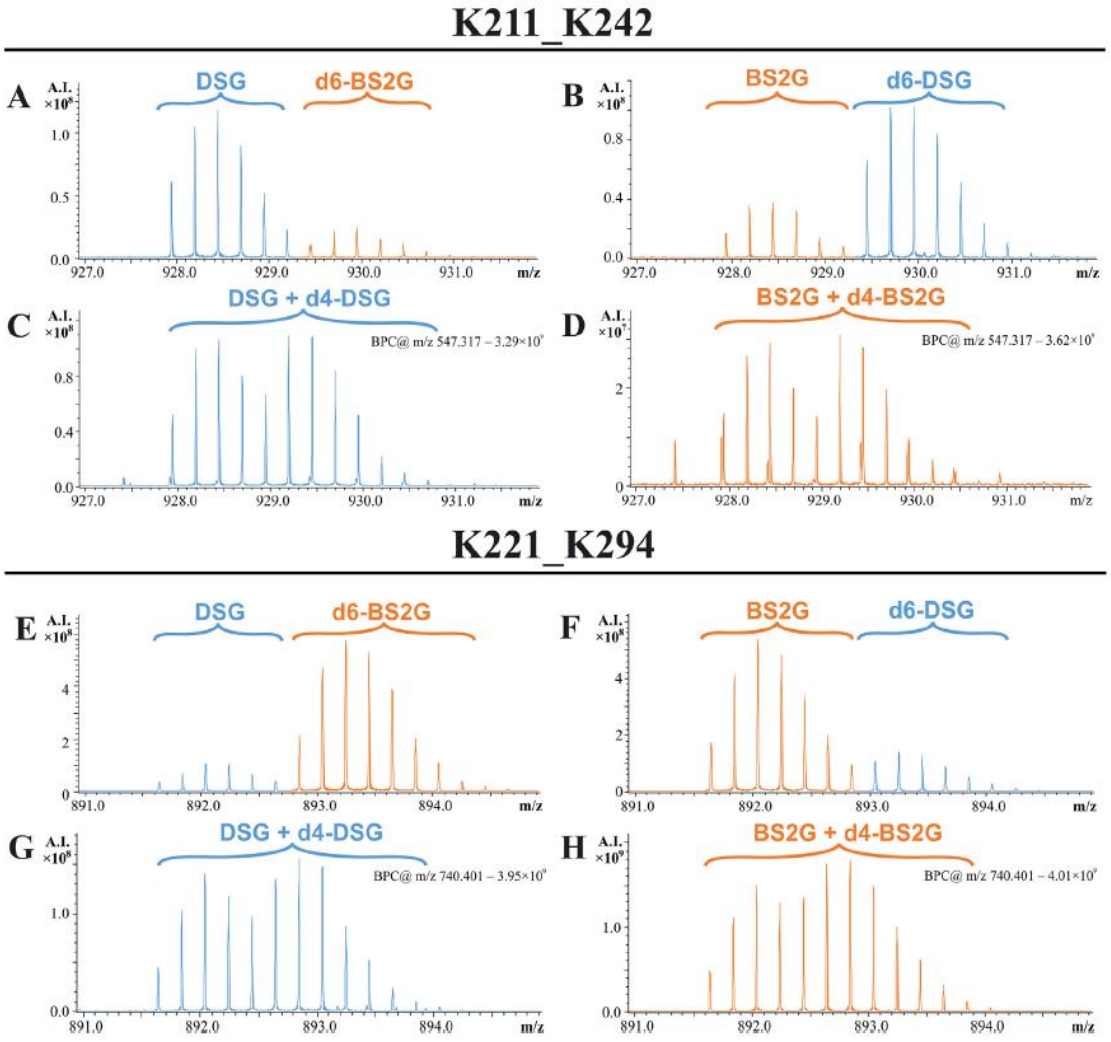


***Figure 54**: Detail of mass spectra corresponding to cross-link K211–K242 (A-D) and K221–K294 (E-H) where isotopically labelled equimolar mixtures composed of DSG/d6-BS2G (A, E) and d6-DSG/BS2G (B, F) were used to track the linker origin. For cross-link K211–K242 (A, B), preference of DSG was observed while cross-link K221–K294 (E, F) was predominantly modified by BS2G. Additionally, inverted ratio patterns were observed (A, B) (E, F) for samples where reciprocal mixtures were used. Control experiments (C, D) and (G, H) resulted in spectra where "doublets" of light/heavy forms were close to the ideal 50:50 ratio.*

Based on the observed phenomenon, quantification of 39 found cross-links was conducted, revealing that half of the found cross-links was preferentially cross-linked by non-polar DSG while only four cross-links preferred polar BS2G.

To explain this phenomenon, all found cross-links were fitted to BSA structural model with a surface representing the polarity of each amino acid. Only one polar and non-polar are shown in **Figure 55**. Not surprisingly, the DSG probe favours hydrophobic regions of proteins molecule while DS2G prefers polar parts.

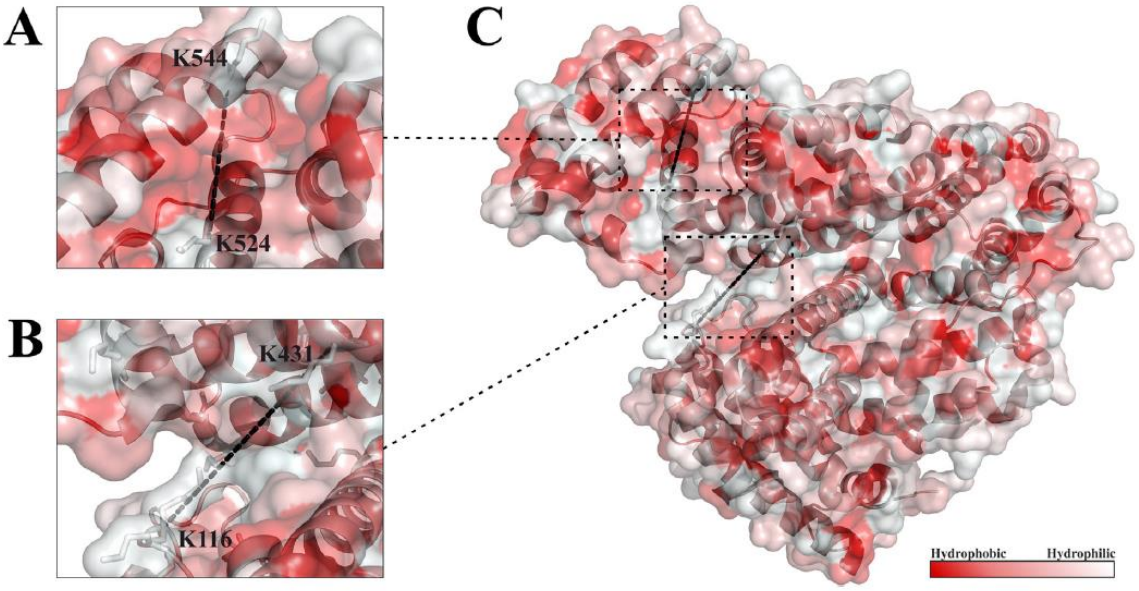


***Figure 55**: Model of BSA (PDB entry: 4F5S) with selected polar and nonpolar cross-links. Based on the colour scheme cross-link K524–K544 (A) is situated near hydrophobic amino acids (red surface colour); thus, preferential modification by DSG was observed in this region. On the contrary, cross-link K116–K431 (B) is located in the polar surface region, making it more accessible for polar BS2G cross-linker.*

Such an assumption and the demand for a high number of constraints keeping the cross-linker concentration as low as possible led to an additional experiment where only light DSG, BS2G, and equimolar mixture DSG/BS2G were used to cross-linking BSA. Cross-linking with DSG and BS2G yield in 47 and 46 found cross-links, respectively. However, the mixture of DSG and BS2G yield 67 unique cross-links within one experiment due to the different preference of the polar and non-polar cross-linkers.

In conclusion, this publication highlighted the role of the cross-linker leaving group, which could significantly affect obtained results even if the cross-linkers have the same spacer length. Additionally, we have shown new insight into sample preparation where the mixture of polar and non-nonpolar cross-linker can be beneficially used to obtain more cross-linking restrain within a single experiment.

4.2.2 Publication 6

Title: Addressing the Molecular Mechanism of Longitudinal Lamin Assembly Using Chimeric Fusions

Since lamins are involved in many cellular processes (chromatin reorganization, transcription, etc.[247]), mutations in the lamin genes are associated with the formation of the unstable nuclear envelope through the incorrect assembly of lamin filaments. This phenomenon is associated with many diseases such as Emery-Dreifuss muscular dystrophy, Hutchinson-Gilford progeria syndrome, collectively known as laminopathies[248]. Although the lamins are intensively studied proteins, the mechanism of filament assembly from monomeric subunits is not fully understood at the molecular level.

For all lamin proteins, the basic building block is a rod-like dimer that forms a parallel coiled-coil (CC) structure. This CC dimeric structure can be divided into smaller domains named head, central α-helical rod, and tail responsible for intermolecular interactions (**Figure 56a**). In previous studies, the predominant longitudinal interaction (so-called A_{CN}) of two dimers with similar orientation was described (**Figure 56b**).

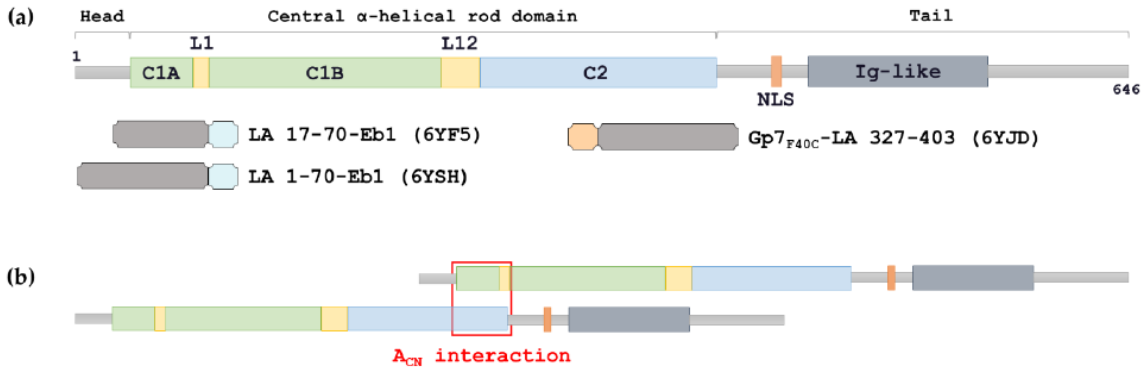


***Figure 56**: a) Schematic representation of the structure of lamin A. C1A, coil1A; L1, linker 1; C1B, coil1B; L12, linker 2; C2, coil2; NLS, nuclear localization signal; Ig-like, immunoglobulin-like domain. Chimeric fragments used here for crystallographic studies are shown below. The capping motifs Eb1 and $Gp7_{F40C}$ are coloured cyan and wheat, respectively. b) Scheme of longitudinal lamin assembly, which is based on the A_{CN} interaction.*

In this publication, the molecular architecture and assembly mechanism of intermediate filaments (IFs) formed by lamin A were studied. X-ray crystallography was the first method of choice to elucidate A_{CN} interaction at the molecular level. However, the crystallization of whole dimers was unsuccessful due to their high flexibility.

Therefore only the short rod domains with stabilizing capping motives responsible for A_{CN} interaction were used for crystallization and X-ray measurement. Specifically, two N-terminal (LA 1-70-Eb1; LA 17-70-Eb1) and one C-terminal ($Gp7_{F40C}$-LA 327-403) fragments were crystalized and structurally solved, representing the head and the tail domain of full-length lamin, respectively.

For the cross-linking experiments, one N-terminal and C-terminal fragments were mixed in a 1:1 ratio to form tetramers through A_{CN} interaction interphase. Subsequently, N-terminal, C-terminal fragments and their complex were subjected to chemical cross-linking by three homobifunctional cross-linkers with primary reactivity against amino groups (DSPU, DSBU and DSG) as well as a zero-length cross-linker EDC with reactivity against amino and carboxy groups. The yields of cross-linking reactions were tracked through SDS-PAGE (**Figure 57**), achieving satisfactory cross-linking products of the individual dimers (2N and 2C) and tetramer (2N2C).

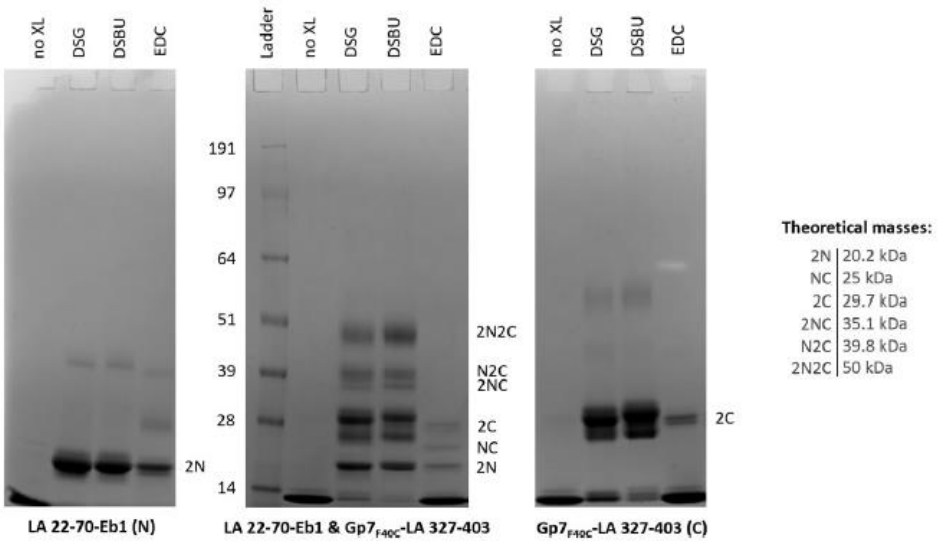


***Figure 57**: Reducing SDS-PAGE analysis of LA 22-70-Eb1, $Gp7_{F40C}$-LA 327-403 and the SEC-purified complex of these two fragments without cross-linking (no XL) and after cross-linking with DSG, DSBU and EDC. The most likely stoichiometry of different bands is given at the right-hand side of each gel, where letters N and C indicate the N- and C-terminal fragments, respectively. Theoretical masses of all bands are given in the side panel.*

Cross-linked samples were then processed by the standard bottom-up protocol followed by data evaluation with Stavrox and Merox software. In total, 51 cross-links were identified for all used cross-linkers. However, only 44 were used for subsequent structural modelling of the 2N2C tetrameric complex (7 cross-links were in the region not resolved in the crystal structure). For molecular modelling, N-terminal and C-terminal dimers were manually

docked upon opening up the interacting ends of each dimer. After optimization steps, an A_{CN} tetramer model with an overlap of 6.5 nm was obtained (**Figure 58a**), satisfying 31 cross-links out of 44 (**Figure 58b**). The 13 remaining cross-links were outside the specified cross-linkers restrains. This effect was additionally explained by the formation of artefacts caused by random collisions of two heterotetramers within the cross-linking reaction.

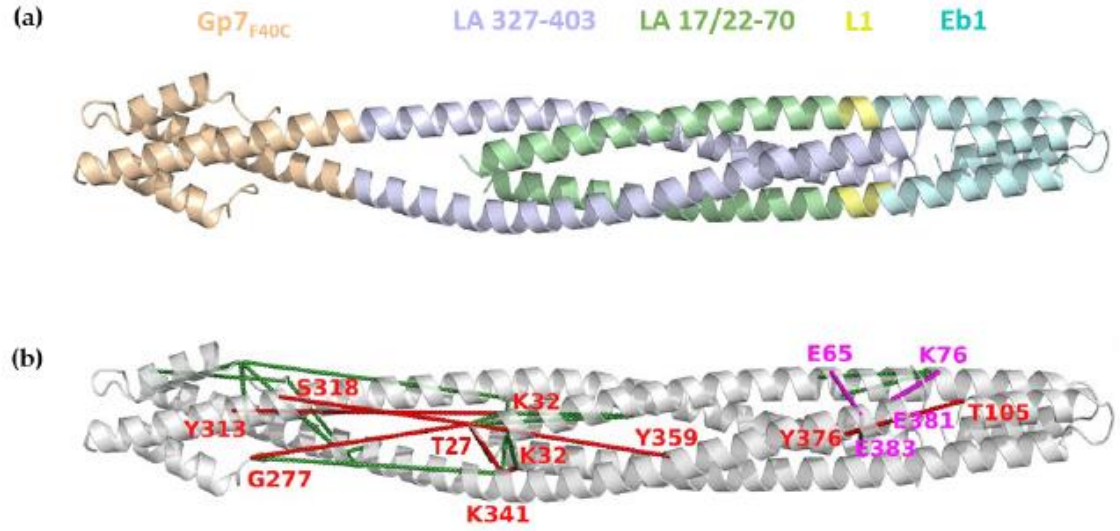


***Figure 58**: Modelling of the A_{CN} tetramer. a) Final molecular model of the A_{CN} tetramer. $Gp7_{F40C}$-LA 327-403 is coloured wheat ($Gp7_{F40C}$) and blue (LA 327-403). LA 17-70-Eb1 is coloured green/yellow (LA 17-70) and cyan (Eb1). b) Final molecular model of the A_{CN} tetramer with all 31 cross-links that were compatible with the model. Interdimer cross-links are coloured green. Tetramer cross-links are labelled and coloured red (DSBU, DSPU and DSG) and magenta (EDC).*

Finally, this publication demonstrated the use of C and N terminal fragments ("mini-lamins") and a combination of X-ray crystallography and chemical cross-linking for the characterization of the longitudinal (A_{CN}) interaction at the molecular level. Thanks to the given structural restrains, molecular modelling provided the tetrameric model with a 6.5 nm overlap of N and C-terminal dimers at A_{CN} interaction interphase. This observation is in excellent agreement with the most recent publication by Makarov *et al.*[249], where the 5-6 nm overlap of full-length lamin A was determined by chemical cross-linking.

5. Summary

This thesis aimed to enhance the capability of structural mass spectrometric methods, especially for pulsed covalent labelling and chemical cross-linking. The results were summarized in three publications and three manuscripts currently being in the peer-review process. The obtained specific results are summarized in the following points:

- Pulsed quench-flow reactors have been designed and constructed to utilize radical FPOP and FFAP labelling methods to determine the surface topology of proteins and protein complexes.
- Firstly, the performance of the quench-flow reactor and the feasibility of the top-down approach for radical labelling were successfully demonstrated on mapping surface accessibility of ubiquitin.
- Subsequently, holo/apomyoglobin structural changes caused by the absence/presence of the heme group were tracked by the FPOP and the FFAP (3 seconds labelling pulse). The independently obtained results from the FPOP and the FFAP radical labelling experiments were mutually in excellent agreement, pinpointing identical protein regions which underwent the most significant structural changes responsible for heme binding.
- Further, the interaction interphase of tetrameric haptoglobin haemoglobin complex was determined at single residue resolution by combining FPOP and FFAP techniques conducted in quench-flow reactors. For this experiments, a novel Bruker timsTOF Pro instrument equipped with IMS and PASEF technology was advantageously used.
- The influence of cross-linker's leaving group polarity (NHS vs sulfo-NHS) was tested. The polar regions of BSA were predominantly cross-linked by polar cross-linker BS2G, while DSG preferred less polar regions. Additionally, results showed that the use of the polar and non-nonpolar cross-linker mixture could be beneficial to obtain more distance constraints within a single experiment.
- Based on the influence of cross-linker polarity, various non-polar cross-linkers were used for cross-linking of "mini-lamins" tetrameric assembly. By a combination of X-ray crystallography and chemical cross-linking, structural model of longitudinal interaction (A_{CN}) of "mini-lamins" was determined with atomic resolution.

LIST OF PUBLICATIONS

Publications included in this thesis:

1. Yassaghi G., Kukačka Z., **Fiala J.**, Kavan D., Halada P., Novák P.

   Fast Photochemical Oxidation of Proteins and Top-Down Mass Spectrometry: The Effect of Collision-Induced Dissociation, Electron-Transfer Dissociation and Electron-Capture Dissociation on Product-Ion Spectra., *Anal. Chem.,* **under review**

2. Loginov D., **Fiala J.**, Chmelík J., Brechlin P., Kruppa G., Novák P.

   Benefits of Ion Mobility Separation and Parallel Accumulation–Serial Fragmentation Technology on timsTOF Pro for the Needs of Fast Photochemical Oxidation of Protein Analysis., *ACS Omega* 2021, 6, 15, 10352–10361. doi:10.1021/acsomega.1c00732

3. Loginov D., **Fiala J.**, Brechlin P., Kruppa G., Novák P.

   Hydroxyl radical footprinting analysis of a human haptoglobin-hemoglobin complex., *Biochim. Biophys. Acta*., **accepted**

4. Fojtík L., **Fiala J.**, Pompach P., Matoušek V., Beier P., Kukačka Z., Novák P.

   Fast Fluoroalkylation of Proteins Uncovers Structure and Dynamics of Biological Macromolecules., *JACS,* **under review**

5. **Fiala J.**, Kukačka Z., Novák P.

   Influence of cross-linker polarity on selectivity towards lysine side chains., *J. Proteomics* 2020, 218, 103716. doi:10.1016/j.jprot.2020.103716

6. Stalmans G., Lilina A. V., Vermeire P. J., **Fiala J.**, Novák P., Strelkov S. V.

   Addressing the Molecular Mechanism of Longitudinal Lamin Assembly Using Chimeric Fusions., *Cells* 2020, 9, 1–21. doi:10.3390/cells9071633

Other publications by author:

7. Slavata L., Chmelík J., Kavan D., Filandrová R., **Fiala J.**, Rosůlek M., Mrázek H.,

   Kukačka Z., Vališ K., Man P., Miller M., McIntyre W., Fabris D., Novák P.

   Ms-based approaches enable the structural characterization of transcription factor/DNA response element complex., *Biomolecules* 2019, 9, 535. doi:10.3390/biom9100535

8. Filandrová R., Vališ K., Černý J., Chmelík J., Slavata L., **Fiala J.**, Rosůlek M., Kavan D., Man P., Chum T., Cebecauer M., Fabris D, Novák P.

Motif orientation matters: Structural characterization of TEAD1 recognition of genomic DNA., *Structure* 2021, 29, 345-356.e8., doi:10.1016/j.str.2020.11.018

BIBLIOGRAPHY

[1] G. J. Mulder, *Bull. des Sci. Phys. Nat. en néerlande* **1838**, *104*, 104–124.

[2] H. Hartley, *Nature* **1951**, *168*, 244.

[3] E. J. Wood, *Nature's Robots: A History of Proteins*, **2004**.

[4] F. Sanger, *Adv. Protein Chem.* **1952**, *7*, 1–67.

[5] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci.* **1951**, *37*, 729–740.

[6] "RCSB PDB - 1SI4: Crystal structure of Human hemoglobin A2 (in R2 state) at 2.2 A resolution," can be found under https://www.rcsb.org/structure/1NQP, *"accesed 06-22-2021"*

[7] A. L. Weber, S. L. Miller, *J. Mol. Evol.* **1981**, *17*, 273–284.

[8] A. Hvidt, K. Linderstrøm-Lang, *BBA - Biochim. Biophys. Acta* **1955**, *16*, 168–169.

[9] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95–99.

[10] E. J. Stollar, D. P. Smith, *Essays Biochem.* **2020**, *64*, 649–680.

[11] S. Choudhuri, *Bioinforma. Beginners* **2014**, 183–207.

[12] S. Gore, E. Sanz García, P. M. S. Hendrickx, A. Gutmanas, J. D. Westbrook, H. Yang, Z. Feng, K. Baskaran, J. M. Berrisford, B. P. Hudson, Y. Ikegawa, N. Kobayashi, C. L. Lawson, S. Mading, L. Mak, A. Mukhopadhyay, T. J. Oldfield, A. Patwardhan, E. Peisach, G. Sahni, M. R. Sekharan, S. Sen, C. Shao, O. S. Smart, E. L. Ulrich, R. Yamashita, M. Quesada, J. Y. Young, H. Nakamura, J. L. Markley, H. M. Berman, S. K. Burley, S. Velankar, G. J. Kleywegt, *Structure* **2017**, *25*, 1916–1927.

[13] Y.-R. Chen, S.-L. Peng, Y.-W. Tsay, in *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Berlin, Heidelberg, **2008**, pp. 204–211.

[14] L. Liu, M. J. White, T. H. MacRae, *Eur. J. Biochem.* **1999**, *262*, 247–257.

[15] H. R. Horton, L. A. Moran, K. G. Scrimgeour, M. D. Perry, J. D. Rawn, *Principles of Biochemistry*, Pearson Prentice Hall, **2006**.

[16] O. Jardetzky, J. F. Lefèvre, *FEBS Lett.* **1994**, *338*, 246–50.

[17] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.

[18] C. Levinthal, *J. Chim. Phys.* **1968**, *65*, 44–45.

[19] A. V. Finkelstein, *Biochem.* **2018**, *83*, S3–S18.

[20] K. A. Dill, H. S. Chan, *Nat. Struct. Biol.* **1997**, *4*, 10–19.

[21] S. W. Englander, L. Mayne, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15873–15880.

[22] N. J. Greenfield, *Nat. Protoc.* **2007**, *1*, 2876–2890.

[23] S. Beychok, *Science.* **1966**, *154*, 1288–1299.

[24] U. B. Gawas, V. K. Mandrekar, M. S. Majik, in *Adv. Biol. Sci. Res.*, Elsevier, **2019**, pp. 69–84.

[25] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662–666.

[26] Y. Shi, *Cell* **2014**, *159*, 995–1014.

[27] "PDB Statistics: Growth of Structures from X-ray Crystallography Experiments Released per Year," can be found under https://www.rcsb.org/stats/growth/growth-xray, *"accesed 06-25-2021"*

[28] J. Holcomb, N. Spellmon, Y. Zhang, M. Doughan, C. Li, Z. Yang, *AIMS Biophys.* **2017**, *4*, 557–575.

[29] N. K. Goto, L. E. Kay, *Curr. Opin. Struct. Biol.* **2000**, *10*, 585–592.

[30] K. Wüthrich, *Nat. Struct. Biol.* **2001**, *8*, 923–925.

[31] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, *64*, 2229–2246.

[32] J. Schraml, J. M. Bellama, S. R. Nattrass, *Two Dimensional NMR Spectroscopy*, **1989**.

[33] E. Z. Eisenmesser, *Science* **2002**, *295*, 1520–1523.

[34] D. F. Gauto, L. F. Estrozi, C. D. Schwieters, G. Effantin, P. Macek, R. Sounier, A. C. Sivertsen, E. Schmidt, R. Kerfah, G. Mas, J. P. Colletier, P. Güntert, A. Favier, G. Schoehn, P. Schanda, J. Boisbouvier, *Nat. Commun.* **2019**, *10*, 1–12.

[35] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, K. H. Downing, *J. Mol. Biol.* **1990**, *213*, 899–929.

[36] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, P. Schultz, *Q. Rev. Biophys.* **1988**, *21*, 129–228.

[37] J. Frank, *Ultramicroscopy* **1975**, *1*, 159–162.

[38] C. Savva, *Biochem. (Lond).* **2019**, *41*, 46–52.

[39] T. Nakane, A. Kotecha, A. Sente, G. McMullan, S. Masiulis, P. M. G. E. Brown, I. T. Grigoras, L. Malinauskaite, T. Malinauskas, J. Miehling, T. Uchański, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. Murshudov, A. R. Aricescu, S. H. W. Scheres, *Nature* **2020**, *587*, 152–156.

[40] M. Yamashita, J. B. Fenn, *J. Phys. Chem.* **1984**, *88*, 4451–4459.

[41] M. Karas, D. Bachmann, F. Hillenkamp, *Anal. Chem.* **1985**, *57*, 2935–2939.

[42] M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, *Int. J. Mass Spectrom. Ion Process.* **1987**, *78*, 53–68.

[43] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo, *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151–153.

[44] J. Griffiths, *Anal. Chem.* **2008**, *80*, 5678–5683.

[45] D. S. Chorev, G. Ben-Nissan, M. Sharon, *Proteomics* **2015**, *15*, 2777–2791.

[46] A. J. R. Heck, *Nat. Methods* **2008**, *5*, 927–933.

[47] A. C. Leney, A. J. R. Heck, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5–13.

[48] V. Katta, B. T. Chait, *J. Am. Chem. Soc.* **1991**, *113*, 8534–8535.

[49] B. Ganem, Y. T. Li, J. D. Henion, *J. Am. Chem. Soc.* **1991**, *113*, 6294–6296.

[50] R. H. H. Van Den Heuvel, A. J. R. Heck, *Curr. Opin. Chem. Biol.* **2004**, *8*, 519–526.

[51] P. Lössl, M. Waterbeemd, A. J. Heck, *EMBO J.* **2016**, *35*, 2634–2657.

[52] S. Deroo, S. J. Hyung, J. Marcoux, Y. Gordiyenko, R. K. Koripella, S. Sanyal, C. V. Robinson, *ACS Chem. Biol.* **2012**, *7*, 1120–1127.

[53] M. W. H. Pinkse, C. S. Maier, J. I. Kim, B. H. Oh, A. J. R. Heck, *J. Mass Spectrom.* **2003**, *38*, 315–320.

[54] J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson, A. J. R. Heck, *Angew. Chemie - Int. Ed.* **2013**, *52*, 4020–4023.

[55] G. Guigas, M. Weiss, *Biochim. Biophys. Acta - Biomembr.* **2016**, *1858*, 2441–2450.

[56] A. L. Duncan, T. Reddy, H. Koldsø, J. Hélie, P. W. Fowler, M. Chavent, M. S. P. Sansom, *Sci. Rep.* **2017**, *7*, 1–15.

[57] L. Konermann, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1827–1835.

[58] Z. Xia, J. B. Degrandchamp, E. R. Williams, *Analyst* **2019**, *144*, 2565–2573.

[59] M. Wilm, M. Mann, *Anal. Chem.* **1996**, *68*, 1–8.

[60] A. El-Faramawy, K. W. M. Siu, B. A. Thomson, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1702–1707.

[61] T. Kenderdine, Z. Xia, E. R. Williams, D. Fabris, *Anal. Chem.* **2018**, *90*, 13541–13548.

[62] A. Ahrens, J. Möhle, M. Hitzemann, S. Zimmermann, *Int. J. Ion Mobil. Spectrom.* **2020**, *23*, 75–81.

[63] B. C. Hauck, W. F. Siems, C. S. Harden, V. M. McHugh, H. H. Hill, *Anal. Chem.* **2018**, *90*, 4578–4584.

[64] R. Cumeras, E. Figueras, C. E. Davis, J. I. Baumbach, I. Gràcia, *Analyst* **2015**, *140*, 1376–1390.

[65] M. J. Cohen, F. W. Karasek, *J. Chromatogr. Sci.* **1970**, *8*, 330–337.

[66] N. Prianichnikov, H. Koch, S. Koch, M. Lubeck, R. Heilig, S. Brehmer, R. Fischer, J. Cox, *Mol. Cell. Proteomics* **2020**, *19*, 1058–1069.

[67] C. G. Vasilopoulou, K. Sulek, A. D. Brunner, N. S. Meitei, U. Schweiger-Hufnagel, S. W. Meyer, A. Barsch, M. Mann, F. Meier, *Nat. Commun.* **2020**, *11*, 1–11.

[68] H. E. Revercomb, E. A. Mason, *Anal. Chem.* **1975**, *47*, 970–983.

[69] W. F. Siems, L. A. Viehland, H. H. Hill, *Analyst* **2016**, *141*, 6396–6407.

[70] V. Gabelica, A. A. Shvartsburg, C. Afonso, P. Barran, J. L. P. Benesch, C. Bleiholder, M. T. Bowers, A. Bilbao, M. F. Bush, J. L. Campbell, I. D. G. Campuzano, T. Causon, B. H. Clowers, C. S. Creaser, E. De Pauw, J. Far, F. Fernandez-Lima, J. C. Fjeldsted, K. Giles, M. Groessl, C. J. Hogan, S. Hann, H. I. Kim, R. T. Kurulugama, J. C. May, J. A. McLean, K. Pagel, K. Richardson, M. E. Ridgeway, F. Rosu, F. Sobott, K. Thalassinos, S. J. Valentine, T. Wyttenbach, *Mass Spectrom. Rev.* **2019**, *38*, 291–320.

[71] R. Salbo, M. F. Bush, H. Naver, I. Campuzano, C. V. Robinson, I. Pettersson, T. J. D. Jørgensen, K. F. Haselmann, *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181–1193.

[72] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V Robinson, *Nat. Protoc.* **2008**, *3*, 1139–1152.

[73] I. Campuzano, M. F. Bush, C. V. Robinson, C. Beaumont, K. Richardson, H. Kim, H. I. Kim, *Anal. Chem.* **2012**, *84*, 1026–1033.

[74] M. F. Bush, I. D. G. Campuzano, C. V. Robinson, *Anal. Chem.* **2012**, *84*, 7124–7130.

[75] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Anal. Chem.* **2010**, *82*, 9557–9565.

[76] J. L. Lippens, S. V. Ranganathan, R. J. D'Esposito, D. Fabris, *Analyst* **2016**, *141*, 4084–4099.

[77] K. Pagel, D. J. Harvey, *Anal. Chem.* **2013**, *85*, 5138–5145.

[78] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinos, R. H. Bateman, M. T. Bowers, J. H. Scrivens, *Int. J. Mass Spectrom.* **2007**, *261*, 1–12.

[79] E. Van Duijn, A. Barendregt, S. Synowsky, C. Versluis, A. J. R. Heck, *J. Am. Chem. Soc.* **2009**, *131*, 1452–1459.

[80] T. W. Knapman, V. L. Morton, N. J. Stonehouse, P. G. Stockley, A. E. Ashcroft,

Rapid Commun. Mass Spectrom. **2010**, *24*, 3033–3042.

[81] E. V. Krylov, E. G. Nazarov, *Int. J. Mass Spectrom.* **2009**, *285*, 149–156.

[82] B. M. Kolakowski, Z. Mester, *Analyst* **2007**, *132*, 842–864.

[83] K. E. Swearingen, R. L. Moritz, *Expert Rev. Proteomics* **2012**, *9*, 505–517.

[84] F. A. Fernandez-Lima, D. A. Kaplan, M. A. Park, *Rev. Sci. Instrum.* **2011**, *82*, 12–15.

[85] M. E. Ridgeway, M. Lubeck, J. Jordens, M. Mann, M. A. Park, *Int. J. Mass Spectrom.* **2018**, *425*, 22–35.

[86] M. Amo-González, S. Pérez, *Anal. Chem.* **2018**, *90*, 6735–6741.

[87] K. Giles, J. Ujma, J. Wildgoose, S. Pringle, K. Richardson, D. Langridge, M. Green, *Anal. Chem.* **2019**, *91*, 8564–8573.

[88] "Valérie Gabelica on Native Mass Spectrometry for Nucleic Acids," can be found under https://www.biotechniques.com/interview/valerie-gabelica-on-native-mass-spectrometry-for-nucleic-acid/, **2019**.

[89] T. Free, *Biotechniques* **2019**, *67*, 204–206.

[90] D. Hewitt, E. Marklund, D. J. Scott, C. V. Robinson, A. J. Borysik, *J. Phys. Chem. B* **2014**, *118*, 8489–8495.

[91] K. B. Shelimov, M. F. Jarrold, *J. Am. Chem. Soc.* **1997**, *119*, 2987–2994.

[92] S. M. Dixit, D. A. Polasky, B. T. Ruotolo, *Curr. Opin. Chem. Biol.* **2018**, *42*, 93–100.

[93] M. Göth, K. Pagel, *Anal. Bioanal. Chem.* **2017**, *409*, 4305–4310.

[94] Y. Tian, L. Han, A. C. Buckner, B. T. Ruotolo, *Anal. Chem.* **2015**, *87*, 11509–11515.

[95] D. W. Woodall, L. W. Henderson, S. A. Raab, K. Honma, D. E. Clemmer, *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 64–72.

[96] A. Pruška, A. Marchand, R. Zenobi, *Angew. Chemie Int. Ed.* **2021**, DOI 10.1002/anie.202016757.

[97] D. W. Woodall, T. J. El-Baba, D. R. Fuller, W. Liu, C. J. Brown, A. Laganowsky, D. H. Russell, D. E. Clemmer, *Anal. Chem.* **2019**, *91*, 6808–6814.

[98] P. Limpikirati, T. Liu, R. W. Vachet, *Methods* **2018**, *144*, 79–93.

[99] A. D. Mirzabekov, A. F. Melnikova, *Mol. Biol. Rep.* **1974**, *1*, 379–384.

[100] D. J. Galas, A. Schmitz, *Nucleic Acids Res.* **1978**, *5*, 3157–3170.

[101] H. Sheshberadaran, L. G. Payne, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1–5.

[102] D. S. Ziemianowicz, J. L. MacCallum, D. C. Schriemer, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 207–216.

[103] L. Wang, M. R. Chance, *Mol. Cell. Proteomics* **2017**, *16*, 706–716.

[104] Y. Pan, L. Konermann, *Analyst* **2010**, *135*, 1191–1200.

[105] V. L. Mendoza, R. W. Vachet, *Mass Spectrom. Rev.* **2009**, *28*, 785–815.

[106] P. Limpikirati, T. Liu, R. W. Vachet, *Methods* **2018**, *144*, 79–93.

[107] N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson, F. W. McLafferty, *J. Am. Chem. Soc.* **1999**, *121*, 806–812.

[108] C. Boone, J. Adamec, *Proteomic Profiling Anal. Chem. Crossroads Second Ed.* **2016**, 175–191.

[109] P. Novak, G. H. Kruppa, M. M. Young, J. Schoeniger, *J. Mass Spectrom.* **2004**, *39*, 322–328.

[110] B. A. Cerda, D. M. Horn, K. Breuker, B. K. Carpenter, F. W. McLafferty, *Eur. J. Mass Spectrom.* **1999**, *5*, 335–338.

[111] J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528–9533.

[112] S. A. Lammert, R. G. Cooks, *J. Am. Soc. Mass Spectrom.* **1991**, *2*, 487–491.

[113] W. D. Bowers, S. S. Delbert, R. L. Hunter, R. T. Mclver, *J. Am. Chem. Soc.* **1984**, *106*, 7288–7289.

[114] J. B. Shaw, W. Li, D. D. Holden, Y. Zhang, J. Griep-Raming, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher, J. S. Brodbelt, *J. Am. Chem. Soc.* **2013**, *135*, 12646–12651.

[115] D. P. Little, J. P. Speir, M. W. Senko, P. B. O'Connor, F. W. McLafferty, *Anal. Chem.* **1994**, *66*, 2809–2815.

[116] F. Meng, A. J. Forbes, L. M. Miller, N. L. Kelleher, *Mass Spectrom. Rev.* **2005**, *24*, 126–134.

[117] T. Liu, P. Limpikirati, R. W. Vachet, *Anal. Chem.* **2019**, *91*, 15248–15254.

[118] K. D. Rand, S. D. Pringle, M. Morris, J. M. Brown, *Anal. Chem.* **2012**, *84*, 1931–1940.

[119] J. J. Englander, D. B. Calhoun, S. W. Englander, *Anal. Biochem.* **1979**, *92*, 517–524.

[120] S. O. Nielsen, *BBA - Biochim. Biophys. Acta* **1960**, *37*, 146–147.

[121] E. R. Blout, C. de Lozé, A. Asadourian, *J. Am. Chem. Soc.* **1961**, *83*, 1895–1900.

[122] A. A. Kossiakoff, *Nature* **1982**, *296*, 713–721.

[123] M. Saunders, A. Wishnia, J. G. Kirkwood, *J. Am. Chem. Soc.* **1957**, *79*, 3289–3290.

[124] M. Saunders, A. Wishnia, *Ann. N. Y. Acad. Sci.* **1958**, *70*, 870–874.

[125] A. Wishnia, *J. Am. Chem. Soc.* **1962**, *84*, 4235–4239.

[126] V. Katta, B. T. Chait, S. Carr, *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214–217.

[127] Z. Zhang, D. L. Smith, *Protein Sci.* **1993**, *2*, 522–531.

[128] A. J. Percy, M. Rey, K. M. Burns, D. C. Schriemer, *Anal. Chim. Acta* **2012**, *721*, 7–21.

[129] A. Hvidt, S. O. Nielsen, *Adv. Protein Chem.* **1966**, *21*, 287–386.

[130] D. D. Weis, T. E. Wales, J. R. Engen, M. Hotchko, L. F. Ten Eyck, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1498–1509.

[131] L. Konermann, J. Pan, Y. H. Liu, *Chem. Soc. Rev.* **2011**, *40*, 1224–1234.

[132] L. M. S Walter Englander, Tobin R Sosnickt, Joan J Englandert, *Curr. Opin. Struct. Biol.* **1996**, *6*, 18–23.

[133] Hui Xiao, Joshua K. Hoerner, Stephen J. Eyles, Andras Dobo, Edward Voigtman, Andre I. Mel'čuk, Igor A. Kaltashov, *Protein Sci.* **2005**, *14*, 543–557.

[134] S. W. Englander, *J Am Soc Mass Spectrom* **2006**, *17*, 1481.

[135] D. L. Smith, Y. Deng, Z. Zhang, *J. Mass Spectrom.* **1997**, *32*, 135–146.

[136] Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins Struct. Funct. Bioinforma.* **1993**, *17*, 75–86.

[137] K. A. Brown, D. J. Wilson, **2017**, *142*, 2874.

[138] S. Arrhenius, *Zeitschrift für Phys. Chemie* **1889**, *4U*, 96–116.

[139] S. W. Englander, M. M. G. Krishna, *Nat. Struct. Biol.* **2001**, *8*, 741–742.

[140] T. R. Keppel, M. E. Jacques, R. W. Young, K. L. Ratzlaff, D. D. Weis, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1472–1476.

[141] L. Konermann, J. Pan, Y.-H. Liu, *This J. is Cite this Chem. Soc. Rev* **2011**, *40*, 1224–1234.

[142] Y. Zhang, D. L. Rempel, J. Zhang, A. K. Sharma, L. M. Mirica, M. L. Gross, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14604–14609.

[143] J. Pan, D. J. Wilson, L. Konermann, *Biochemistry* **2005**, *44*, 8627–8633.

[144] I. Oganesyan, C. Lento, D. J. Wilson, *Methods* **2018**, *144*, 27–42.

[145] G. R. Masson, J. E. Burke, N. G. Ahn, G. S. Anand, C. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. Faber, R. Garlish, P. R. Griffin, M. L. Gross, M. Guttman, Y. Hamuro, A. J. R. Heck, D. Houde, R. E. Iacob, T. J. D. Jørgensen, I. A. Kaltashov, J. P. Klinman, L. Konermann, P. Man, L. Mayne, B. D. Pascal, D. Reichmann, M. Skehel, J. Snijder, T. S. Strutzenberg, E. S. Underbakke, C. Wagner, T. E. Wales, B. T. Walters, D. D. Weis, D. J. Wilson, P. L. Wintrode, Z. Zhang, J.

Zheng, D. C. Schriemer, K. D. Rand, *Nat. Methods* **2019**, *16*, 595–602.
[146] T. R. Keppel, D. D. Weis, *Anal. Chem.* **2013**, *85*, 5161–5168.
[147] J. Pan, J. Han, C. H. Borchers, L. Konermann, *Anal. Chem.* **2010**, *82*, 8591–8597.
[148] D. A. Simmons, S. D. Dunn, L. Konermann, *Biochemistry* **2003**, *42*, 5896–5905.
[149] T. Rob, P. K. Gill, D. Golemi-Kotra, D. J. Wilson, *Lab Chip* **2013**, *13*, 2528–2532.
[150] G. T. Hermanson, *Bioconjugate Techniques: Third Edition*, Elsevier Inc., **2013**.
[151] D. Suckau, M. Mak, M. Przybylski, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5630–5634.
[152] H. Pauly, *Zeitschrift für Physiol. Chemie* **1904**, *42*, 508–518.
[153] B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **1998**, *281*, 269–272.
[154] A. Sinz, *Anal. Bioanal. Chem.* **2010**, *397*, 3433–3440.
[155] J. F. Tait, C. Smith, Z. Levashova, B. Patel, F. G. Blankenberg, J. L. Vanderheyden, *J. Nucl. Med.* **2006**, *47*, 1546–1583.
[156] E. Engvall, P. Perlmann, *J. Immunol.* **1972**, *109*.
[157] D. R. Lide, *CRC Handbook of Chemistry and Physics*, CRC Press, Bosa Roca, **2004**.
[158] J. M. Hooker, A. P. Esser-Kahn, M. B. Francis, *J. Am. Chem. Soc.* **2006**, *128*, 15558–15559.
[159] N. Jentoft, D. G. Dearborn, *J. Biol. Chem.* **1979**, *254*, 4359–4365.
[160] K. Tanaka, Y. Fujii, K. Fukase, *ChemBioChem* **2008**, *9*, 2392–2397.
[161] E. Baslé, N. Joubert, M. Pucheault, *Chem. Biol.* **2010**, *17*, 213–227.
[162] J. S. Richardson, D. C. Richardson, in *Predict. Protein Struct. Princ. Protein Conform.*, Springer US, **1989**, pp. 1–98.
[163] J. V. Comasseto, A. S. Guarezemini, *Sulfur, Selenium, and Tellurium* **2008**, *02138*, 1.
[164] T. P. Begley, *Wiley Encyclopedia of Chemical Biology*, John Wiley & Sons, Inc., **2007**.
[165] N. Nakajima, Y. Ikada, *Bioconjug. Chem.* **1995**, *6*, 123–130.
[166] M. A. Gilles, A. Q. Hudson, C. L. Borders, *Anal. Biochem.* **1990**, *184*, 244–248.
[167] A. Wormall, *J. Exp. Med.* **1930**, *51*, 295–317.
[168] T. J. Tsomides, H. N. Eisen, *Anal. Biochem.* **1993**, *210*, 129–135.
[169] R. A. G. Smith, J. R. Knowles, *J. Chem. Soc. Perkin Trans. 2* **1975**, 686–694.
[170] M. Suchanek, A. Radzikowska, C. Thiele, *Nat. Methods* **2005**, *2*, 261–267.
[171] D. M. Hambly, M. L. Gross, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057–2063.
[172] S. Vahidi, L. Konermann, *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1156–1164.

[173] D. Rozbeský, M. Rosůlek, Z. Kukačka, J. Chmelík, P. Man, P. Novák, *Anal. Chem.* **2018**, *90*, 1104–1113.

[174] K. Takamoto, M. R. Chance, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 251–276.

[175] G. V. Buxton, C. L. Greenstock, W. P. Helman, A. B. Ross, *J. Phys. Chem. Ref. Data* **1988**, *17*, 513–886.

[176] W. M. Garrison, *Chem. Rev.* **1987**, *87*, 381–398.

[177] G. Xu, M. R. Chance, *Chem. Rev.* **2007**, *107*, 3514–3543.

[178] H. J. H. Fenton, *J. Chem. Soc. Trans.* **1894**, *65*, 899–910.

[179] F. Haber, J. Weiss, *Naturwissenschaften* **1932**, *20*, 948–950.

[180] F. Haber, J. Weiss, J. O. Seph, W. Eiss, *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **1934**, *147*, 332–351.

[181] X. R. Liu, M. M. Zhang, M. L. Gross, *Chem. Rev.* **2020**, *120*, 4355–4454.

[182] J. S. Sharp, J. M. Becker, R. L. Hettich, *Anal. Biochem.* **2003**, *313*, 216–225.

[183] B. C. Gau, J. S. Sharp, D. L. Rempel, M. L. Gross, *Anal. Chem.* **2009**, *81*, 6563–6571.

[184] B. Sclavi, S. Woodson, M. Sullivan, M. R. Chance, M. Brenowitz, *J. Mol. Biol.* **1997**, *266*, 144–159.

[185] J. L. Weeks, M. S. Matheson, *J. Am. Chem. Soc.* **1956**, *78*, 1273–1278.

[186] J. S. Sharp, J. M. Becker, R. L. Hettich, *Anal. Chem.* **2004**, *76*, 672–683.

[187] T. T. Aye, T. Y. Low, S. K. Sze, *Anal. Chem.* **2005**, *77*, 5814–5822.

[188] L. Konermann, B. B. Stocks, T. Czarny, *Anal. Chem.* **2010**, *82*, 6667–6674.

[189] Y. Zhang, D. L. Rempel, H. Zhang, M. L. Gross, *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 526–529.

[190] K. K. J. Chan, D. O'Hagan, in *Methods Enzymol.*, Academic Press Inc., **2012**, pp. 219–235.

[191] H. J. Emeléus, R. N. Haszeldine, *J. Chem. Soc.* **1949**, 2948–2952.

[192] U. Teruo, I. Sumi, *Tetrahedron Lett.* **1990**, *31*, 3579–3582.

[193] P. Eisenberger, S. Gischig, A. Togni, *Chem. - A Eur. J.* **2006**, *12*, 2579–2586.

[194] S. Mizuta, S. Verhoog, K. M. Engle, T. Khotavivattana, M. O'Duill, K. Wheelhouse, G. Rassias, M. Médebielle, V. Gouverneur, *J. Am. Chem. Soc.* **2013**, *135*, 2505–2508.

[195] S. Chaabouni, F. Simonet, A. François, S. Abid, C. Galaup, S. Chassaing, *European J. Org. Chem.* **2017**, *2017*, 271–277.

[196] J. Charpentier, N. Früh, A. Togni, *Chem. Rev.* **2015**, *115*, 650–682.

[197] B. R. Langlois, E. Laurent, N. Roidot, *Tetrahedron Lett.* **1991**, *32*, 7525–7528.

[198] Y. Ji, T. Brueckl, R. D. Baxter, Y. Fujiwara, I. B. Seiple, S. Su, D. G. Blackmond, P. S. Baran, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14411–14415.

[199] Y. Fujiwara, J. A. Dixon, F. O'hara, E. D. Funder, D. D. Dixon, R. A. Rodriguez, R. D. Baxter, B. Herlé, N. Sach, M. R. Collins, Y. Ishihara, P. S. Baran, *Nature* **2012**, *492*, 95–99.

[200] M. Cheng, B. Zhang, W. Cui, M. L. Gross, *Angew. Chemie - Int. Ed.* **2017**, *56*, 14007–14010.

[201] D. Eliezer, P. E. Wright, *J. Mol. Biol.* **1996**, *263*, 531–538.

[202] J. Pan, J. Han, C. H. Borchers, L. Konermann, *J. Am. Chem. Soc.* **2009**, *131*, 12801–12808.

[203] S. Caponea, I. Kieltschb, O. Flögela, G. Lelaisa, A. Togni, D. Seebach, *Helv. Chim. Acta* **2008**, *91*, 2035–2056.

[204] K. Rahimidashaghoul, I. Klimánková, M. Hubálek, M. Korecký, M. Chvojka, D. Pokorný, V. Matoušek, L. Fojtík, D. Kavan, Z. Kukačka, P. Novák, P. Beier, *Chem. - A Eur. J.* **2019**, *25*, 15779–15785.

[205] B. C. Gau, H. Chen, Y. Zhang, M. L. Gross, *Anal. Chem.* **2010**, *82*, 7821–7827.

[206] M. M. Zhang, D. L. Rempel, M. L. Gross, *Free Radic. Biol. Med.* **2019**, *131*, 126–132.

[207] F. M. Richards, R. Lamed, R. Wynn, D. Patel, G. Olack, *Protein Sci.* **2004**, *13*, 1706–1706.

[208] T. Ly, R. R. Julian, *J. Am. Chem. Soc.* **2008**, *130*, 351–358.

[209] V. H. Zahn, J. Meienhofer, *Die Makromol. Chemie* **1958**, *26*, 153–166.

[210] T. B. Farmer, R. M. Caprioli, *Biol. Mass Spectrom.* **1991**, *20*, 796–800.

[211] M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, G. Dollinger, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5802–5806.

[212] J. D. Chavez, J. E. Bruce, *Curr. Opin. Chem. Biol.* **2019**, *48*, 8–18.

[213] A. Sinz, *Mass Spectrom. Rev.* **2006**, *25*, 663–682.

[214] J. Pi, L. Sael, *Int. J. Mol. Sci.* **2013**, *14*, 20635–20657.

[215] B. Schilling, R. H. Row, B. W. Gibson, X. Guo, M. M. Young, *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834–850.

[216] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, *Mol. Cell. Proteomics* **2010**, *9*, 1634–1649.

[217] P. D. Bragg, C. Hou, *Arch. Biochem. Biophys.* **1975**, *167*, 311–321.

[218] A. J. Lomant, G. Fairbanks, *J. Mol. Biol.* **1976**, *104*, 243–261.

[219] J. V. Staros, *Biochemistry* **1982**, *21*, 3950–3955.

[220] P. S. R. Anjaneyulu, J. V. Staros, *Int. J. Pept. Protein Res.* **1987**, *30*, 117–124.

[221] M. D. Leavell, P. Novak, C. R. Behrens, J. S. Schoeniger, G. H. Kruppa, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1604–1611.

[222] S. Kalkhof, A. Sinz, *Anal. Bioanal. Chem.* **2008**, *392*, 1–8.

[223] B. Steigenberger, P. Albanese, A. J. R. Heck, R. A. Scheltema, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 196–206.

[224] F. J. O'Reilly, J. Rappsilber, *Nat. Struct. Mol. Biol.* **2018**, *25*, 1000–1008.

[225] D. R. Müller, P. Schindler, H. Towbin, U. Wirth, H. Voshol, S. Hoving, M. O. Steinmetz, *Anal. Chem.* **2001**, *73*, 1927–1934.

[226] E. V. Petrotchenko, V. K. Olkhovik, C. H. Borchers, *Mol. Cell. Proteomics* **2005**, *4*, 1167–1179.

[227] C. Yu, L. Huang, *Anal. Chem.* **2018**, *90*, 144–165.

[228] L. Fischer, Z. A. Chen, J. Rappsilber, *J. Proteomics* **2013**, *88*, 120–128.

[229] R. Zhang, C. S. Sioma, S. Wang, F. E. Regnier, *Anal. Chem.* **2001**, *73*, 5142–5149.

[230] A. Kao, C. Chiu, D. Vellucci, Y. Yang, V. R. Patel, S. Guan, A. Randall, P. Baldi, S. D. Rychnovsky, L. Huang, *Mol. Cell. Proteomics* **2011**, *10*, M110.002170.

[231] M. Q. Müller, F. Dreiocker, C. H. Ihling, M. Schäfer, A. Sinz, *Anal. Chem.* **2010**, *82*, 6958–6968.

[232] C. H. Ihling, P. Springorum, C. Iacobucci, C. Hage, M. Götze, M. Schäfer, A. Sinz, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 183–189.

[233] P. Novák, G. H. Kruppa, *Eur. J. Mass Spectrom.* **2008**, *14*, 355–365.

[234] A. Leitner, L. A. Joachimiak, P. Unverdorben, T. Walzthoeni, J. Frydman, F. Förster, R. Aebersold, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9455–9460.

[235] M. Scalabrin, S. M. Dixit, M. M. Makshood, C. E. Krzemien, D. Fabris, *Methods* **2018**, *144*, 64–78.

[236] A. F. Gomes, F. C. Gozzo, *J. Mass Spectrom.* **2010**, *45*, 892–899.

[237] K. Straub, D. Kanne, H. Rapoport, S. T. Isaacs, K. Straub, *Q. Rev. Biophys.* **1984**, *17*, 1–44.

[238] A. J. Deans, S. C. West, *Nat. Rev. Cancer* **2011**, *11*, 467–480.

[239] A. Sinz, *Angew. Chemie - Int. Ed.* **2018**, *57*, 6390–6396.

[240] L. Konermann, B. B. Stocks, T. Czarny, in *Anal. Chem.*, **2010**, pp. 6667–6674.

[241] S. Vahidi, B. B. Stocks, Y. Liaghati-Mobarhan, L. Konermann, *Anal. Chem.* **2012**, *84*, 9124–9130.

[242] R. Maurus, C. M. Overall, R. Bogumil, Y. Luo, A. G. Mauk, M. Smith, G. D. Brayer, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1997**, *1341*, 1–13.

[243] "timsTOF Pro | Bruker," can be found under https://www.bruker.com/en/products-and-solutions/mass-spectrometry/timstof/timstof-pro.html,*"accesed 06-20-2021"***.**

[244] C. B. F. Andersen, M. Torvund-Jensen, M. J. Nielsen, C. L. P. De Oliveira, H. P. Hersleth, N. H. Andersen, J. S. Pedersen, G. R. Andersen, S. K. Moestrup, *Nature* **2012**, *489*, 456–459.

[245] K. Stødkilde, M. Torvund-Jensen, S. K. Moestrup, C. B. F. Andersen, *Nat. Commun.* **2014**, *5*, 1–8.

[246] C. Nantasenamat, V. Prachayasittikul, L. Bulow, *PLoS One* **2013**, *8*, e62996.

[247] T. Shimi, K. Pfleghaar, S. I. Kojima, C. G. Pack, I. Solovei, A. E. Goldman, S. A. Adam, D. K. Shumaker, M. Kinjo, T. Cremer, R. D. Goldman, *Genes Dev.* **2008**, *22*, 3409–3421.

[248] S. mi Kang, M. H. Yoon, B. J. Park, *BMB Rep.* **2018**, *51*, 327–337.

[249] A. A. Makarov, J. Zou, D. R. Houston, C. Spanos, A. S. Solovyova, C. Cardenal-Peralta, J. Rappsilber, E. C. Schirmer, *Nat. Commun.* **2019**, *10*, 3056.

ATTACHED PUBLICATIONS

ARTICLE I

Yassaghi G., Kukačka Z., Fiala J., Kavan D., Halada P., Novák P.

Fast Photochemical Oxidation of Proteins and Top-Down Mass Spectrometry: The Effect of Collision-Induced Dissociation, Electron-Transfer Dissociation and Electron-Capture Dissociation on Product-Ion Spectra.

Anal. Chem., **under review**

My contribution: Development and testing of apparatus for FPOP experiment.

Fast Photochemical Oxidation of Proteins and Top-Down Mass Spectrometry: The Effect of Collision-Induced Dissociation, Electron-Transfer Dissociation and Electron-Capture Dissociation on Product-Ion Spectra

Ghazaleh Yassaghi[†,§], Zdeněk Kukačka[†,§], Jan Fiala[†,‡], Daniel Kavan[†,‡], Petr Halada[†] and Petr Novák[*,†,‡]

[†]Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic; [‡]Faculty of Science, Charles University in Prague, Prague, Czech Republic

KEYWORDS: Fast photochemical oxidation of proteins, top-down, collision-induced dissociation, electron-transfer dissociation, electron-capture dissociation, mass spectrometry.

ABSTRACT: Fast photochemical oxidation of proteins (FPOP) hydroxyl radical footprinting is a structural mass spectrometry method providing information about protein conformation, structural dynamics, and interactions. Bottom-up mass spectrometry is a standard method to analyze FPOP samples, however protease digestion of all proteins in the sample results in effective mixing of data from various subpopulations of sample proteins with varying degrees of modification. We believe that to assess FPOP data correctly, only singly oxidized protein populations where the spatial distribution of oxidation truly corresponds to initial protein solvent accessibility and residue reactivity unaffected by another previous oxidative modification should be considered for analysis. Therefore, we employ top-down mass spectrometry and test collision-induced dissociation (CID), electron-transfer dissociation (ETD), and electron-capture dissociation (ECD) fragmentation techniques, along with their multi continuous accumulation of selected ions (CASI) variants, to find a suitable method to isolate, fragment, and evaluate singly oxidized subpopulation of FPOP labeled ubiquitin as a single protein model and apo/holo-myoglobin for visualization of structural changes induced by cofactor absence.

Understanding biological processes requires describing the exact structure, interactions, and structural dynamics of involved molecules. With the advent of structural mass spectrometry, one way to obtain structural information about proteins is protein footprinting, which can be used to elucidate solvent accessibility or structural changes in proteins based on the ratio of labeled and unlabeled peptides from different parts of protein or under different conditions[1,2]. In addition to common non-covalent labeling such as hydrogen-deuterium exchange (HDX), covalent labeling by reactive radicals can also be used effectively. While there is a broad selection of radicals[3–5] that could be used as a probe with different reactivity, hydroxyl radicals were found to be efficient, relatively easy to generate, and able to oxidize and thus label side chains of many different amino acids results in good spatial resolution. Several ways to generate hydroxyl radicals were developed for labeling purposes, for example, synchrotron radiolysis[6], Fenton reaction[7] or photolysis of hydrogen peroxide. The latter is used in Fast photochemical oxidation of proteins (FPOP), where an excimer laser generates hydroxyl radicals in flow setup, used to label the molecules of interest for a short time window[4,8]. Since the inception of the method, several other radicals, namely sulfate radical anion[9], carbene diradical[5], carbonate radical anion[10], and trifluoromethyl radical[11], have also been used successfully. The label's covalent nature means the information is stable, unlike deuteration in HDX, resulting in simple sample handling.

Because the oxidative modification is occurring on a single residue, the method has potentially higher resolution, but it depends on a suitable amino acid composition of the protein. The oxidative modification by hydroxyl radical usually results in the +16 Da mass shift of the oxidized amino acids, even though different mass shifts appear depending on residue modified[12]. Various amino acid residues have different reactivity toward •OH which was previously experimentally found to be in following order: Cys > Met > Trp > Tyr > Phe > His > Leu ~ Ile > Arg ~ Lys ~ Val > Ser ~ Thr ~ Pro > Gln ~ Glu > Asp ~ Asn > Ala > Gly[13,14]. A recent study of hydroxyl radical reaction pathways for 13 amino acids by Gross *et al.* classified residues into three classes according to their reactivity to hydroxyl radicals in the FPOP set up[15].

The usual way of analyzing FPOP labeled proteins is bottom-up mass spectrometric analysis. A protease digests labeled proteins to yield peptides, which are then subjected to LC-MS/MS analysis. The MS/MS fragmentation enables to pinpoint oxidized residue's location, and a suitable LC method is used to separate isobaric peptides differing only by the position of oxidative modification. While this method is relatively efficient and can yield good sequence coverage if suitable protease is used for protein digestion, pinpointing a specifically modified residue can still be challenging. Moreover, the analyzed peptides originate from a sum of all proteoforms present in the sample. Thus, there is an intrinsic variability hidden in the data. "Overmodified" proteins can have their structure altered by the initial modification in a way that could increase or decrease an affinity to additional modification in different parts of the protein and thus add artifacts to the data[16]. Theoretically, when labeling a protein, the most reliable information would come out of singly modified proteoforms. In our case, one hydroxyl radical attacks unlabeled protein at all accessible and modifiable residues. The resulting statistical distribution of a single modification at different residues would give a reliable picture of the initial protein structure and dynamics. In this regard, top-down mass spectrometry is more suitable than a bottom-up approach[17]. In the top-down approach, intact protein mass is measured, and ions of specific m/z values can be selected for fragmentation. It is, therefore, possible to only analyze proteoforms with a specific extent of stable covalent modification[18–20].

In this work, we describe the use of top-down mass spectrometry to study FPOP labeled ubiquitin and to monitor conformational changes in labeled myoglobin in the presence/absence of heme group using 248 nm KrF excimer laser for hydroxyl radical generation. Collision-induced dissociation (CID), electron-transfer dissociation (ETD)[21], electron-capture dissociation (ECD)[22], and novel multi continuous accumulation of selected ions (CASI) CID and ECD fragmentation techniques were used to analyze isolated singly oxidized proteoforms of ubiquitin or myoglobin to obtain more detailed information about position and extent of modification. We also use a new home-built software tool to quantify the extent of modification for all observed ions providing an accessible and feasible way to analyze and visualize top-down MS/MS data for the oxidized proteins. Ms2links[23] algorithm was applied for fragment annotation.

RESULTS AND DISCUSSION

The first experiments used well-known and extensively studied ubiquitin as a model protein. We submitted the ubiquitin to FPOP, with hydroxyl radicals generated by photolysis of low concentration (0.03%) of hydrogen peroxide using KrF excimer laser, followed by top-down analysis, to obtain detailed residue-level information. **Figure 1** shows the mass spectrum of both non-oxidized (**Figure 1a**) and oxidized ubiquitin (**Figure 1b**) with charge states ranging from 6+ to 12+. The most intense signal corresponding to the 10+ charge state was chosen to display the relative extent of the protein's oxidation. Without laser irradiation, only the peak of non-oxidized ubiquitin together with its sodium adduct was presented (**Figure 1a**). When irradiated (**Figure 1b**), singly (+15.995 Da), doubly (+31.990 Da), and triply oxidized ubiquitin (+47.985 Da) appeared.

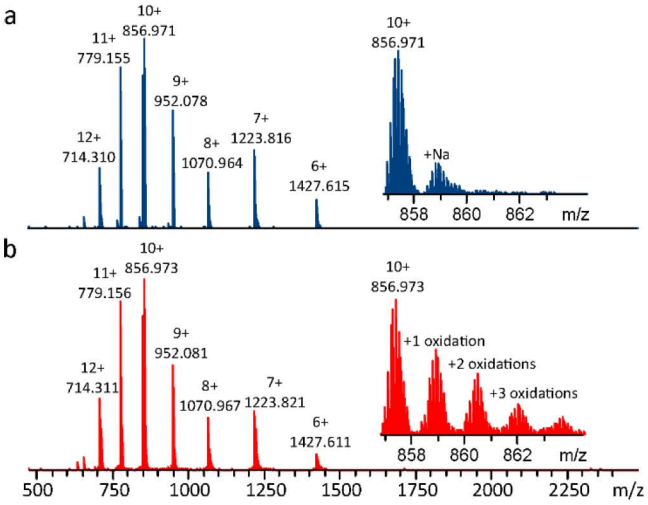


Figure 1: Full mass spectra of the non-oxidized (a) and oxidized (b) ubiquitin. The inset shows the mass spectra corresponding to the 10+ charge state of respective proteins.

Comparison of both spectra illustrates the presence of singly, doubly, and triply oxidized ubiquitin in the sample subjected to FPOP.

In order to identify oxidized residues, singly oxidized ubiquitin of 10+ charge state was isolated and subjected to CID, ETD, and ECD fragmentations. Also, singly oxidized ubiquitin in four different charge states was isolated and fragmented using multi CASI/CID and multi CASI/ECD to obtain more informative MS/MS spectra.

The first fragmentation method tested was CID, which generated b and y ions. CID fragmentation spectrum of singly oxidized ubiquitin (10+ charge state) indicates that both b and y ions are also present in oxidized forms with a mass shift of +16 Da (**Figure 2a**). The quantification of FPOP oxidation derived from b-, y-type ions, and internal fragments is shown in Figure 2b for ions originating from the 10+ charge state precursor and Figure S1 ions of 9+, 10+, 11+, and 12+ charge state spectra. The extent of oxidation was calculated using the equation (**eq.1**) described in the method section. The percentage of fragment ion intensities found in oxidized forms was calculated for all detected ions.

As an example, peaks at m/z 260.11 and m/z 276.10 correspond to b2 and b2+O, respectively. Since the b2 ion is carrying methionine and glutamine, we can safely assume that the approximately 10% calculated oxidation rate belongs to methionine, since glutamine has very low reactivity to hydroxyl radicals, compared to methionine. A jump up in oxidation rate follows in b3, revealing that an additional roughly 12% of oxidation rate belongs to Ile3. A stable level of oxidation up to b36 ion reveals that no additional oxidation occurs in that region. Unfortunately, for larger fragments uniquely covering sequence, which contains more potentially reactive residues, it is impossible to localize the site of oxidation with such certainty. However, known reactivity of residues toward •OH radicals can be used to make an educated guess about the most probable site of oxidation, but higher number of fragments would be needed to achieve better spatial resolution. Such phenomena could be found in y-type ion series, where first ion y12 covers four leucines and one histidine, all potentially oxidizable. The total extent of oxidation in y12 will thus be a sum of oxidation on all five reactive residues depending on which one is more solvent exposed or intrinsically reactive.

The obtained data showed, that sequence coverage provided by b and y ion series is insufficient to obtain reasonable spatial resolution and the internal fragment ions must be taken into the account (**Figure 2b**). The minor oxidation of b19-24 fragment is observed with no further increase toward Asp39 (b19-39) suggesting oxidation of Pro19 or Ile23 for the entire sequence region. An increase in modification of b19-52 ion detects the oxidation of Phe45 since it is the most susceptible residue to oxidation within the 40–52 sequence region. The identical oxidation extent for b19-52 and b37-58 internal ions excludes the oxidation of the sequence region 53–58.

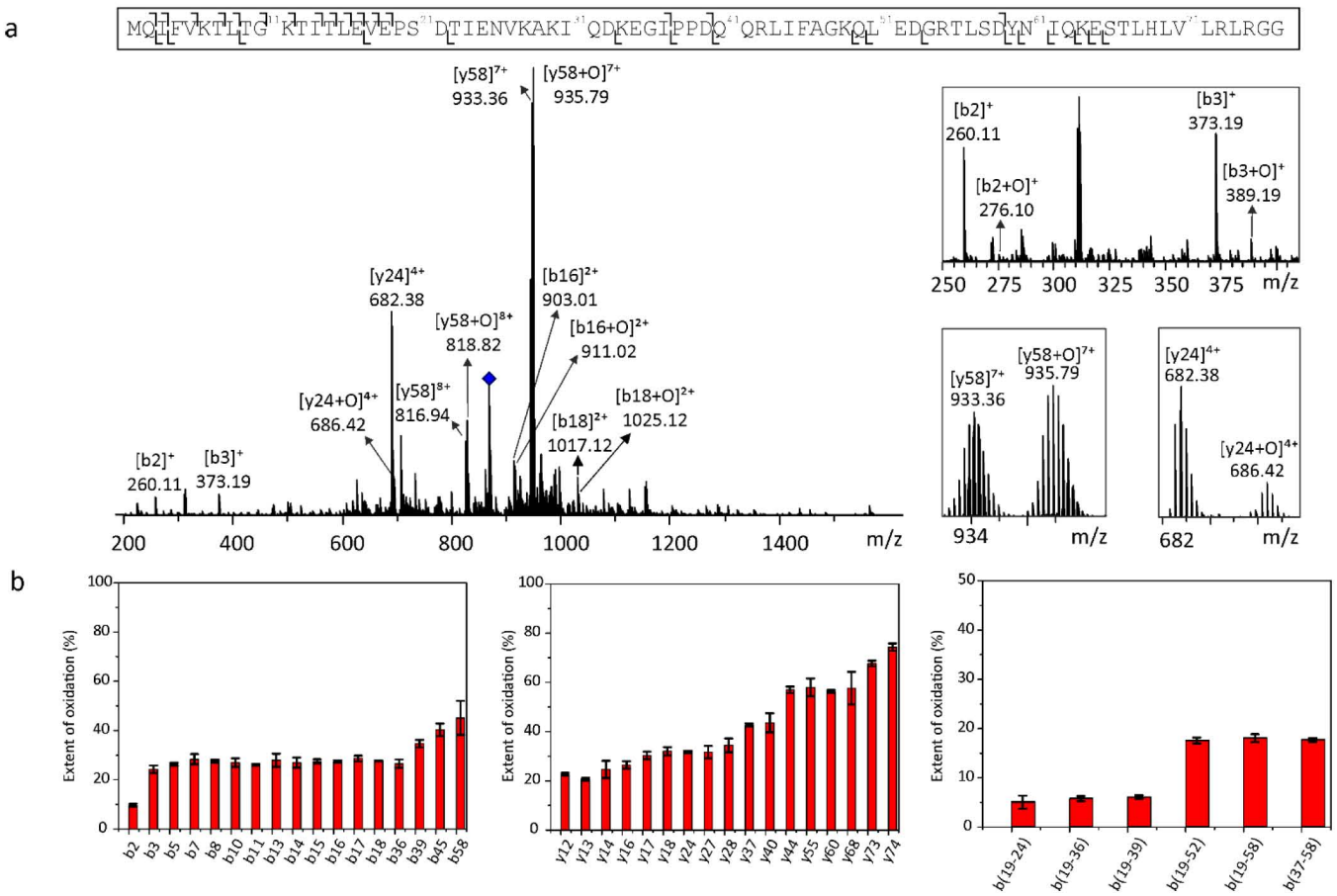


Figure 2: CID MS/MS spectra of FPOP labeled ubiquitin. a) Full MS/MS spectrum obtained upon isolation of 10+ charge state of singly oxidized ubiquitin (m/z 858.4), indicated with a blue diamond. CID MS/MS spectrum of the m/z range 250–415 and spectra of y24 and y58 ions are in boxes on the right. **b)** Extent of FPOP oxidation in detected b and y ions. Red bars highlight the oxidized fragment ions. Error bars are showing the standard deviations of three independent MS measurements.

Although internal fragment ions improve the spatial resolution, CID fragmentation does not allow unambiguous assignment of modifications on a residue level. Therefore, ETD and ECD fragmentation techniques were employed[22], which should, in theory, yield higher sequence coverage[24]. For ECD and ETD, the quadrupole was set to isolate desired ubiquitin ions. 10+ charged singly oxidized ubiquitin was fragmented, which was consistent with the CID experiment. Both ETD and ECD generated c and z˙ ion types, and the observed fragment ions are shown in **Figure S2** and **Figure 3**. ETD-induced fragmentation was inefficient, with the sequence coverage even worse than CID, although the extents of oxidation were relatively consistent with the CID experiment (**Figure S2**). Compared to CID or ETD, ECD fragmentation yielded a much higher number of fragments, resulting in significantly higher sequence coverage and resolution. The oxidative modification can be assigned to specific residues (**Figure 3**). The oxidation profile is similar to those obtained by CID or ETD. The highest increase of oxidation can be seen for c2 fragment, verifying significant oxidation of Met1, and then for c4 fragment suggesting that Phe4 is readily oxidized as well. Increase of oxidation extent for c ions in the region 38-59 is in a nice agreement with CID and ETD data. A clear jump up in oxidation between c58 and c59 can be assigned unambiguously to a single residue Tyr59. In comparison to y ions from CID, z˙ ions obtained by ECD offer better coverage of last 12 amino acid residues previously seen as single fragment and show individual increases in oxidation belonging to reactive residues Leu73, Leu71, Leu69, and Leu67. Additionally, different charge states of singly oxidized ubiquitin resulted in somewhat different fragmentation behavior and, more importantly, different extents of oxidation in CID. Comparison of oxidation extent analysis on fragments obtained from different charge states of ubiquitin (from 9+ to 12+) is shown in **Figure 4** and **Figure S1**.

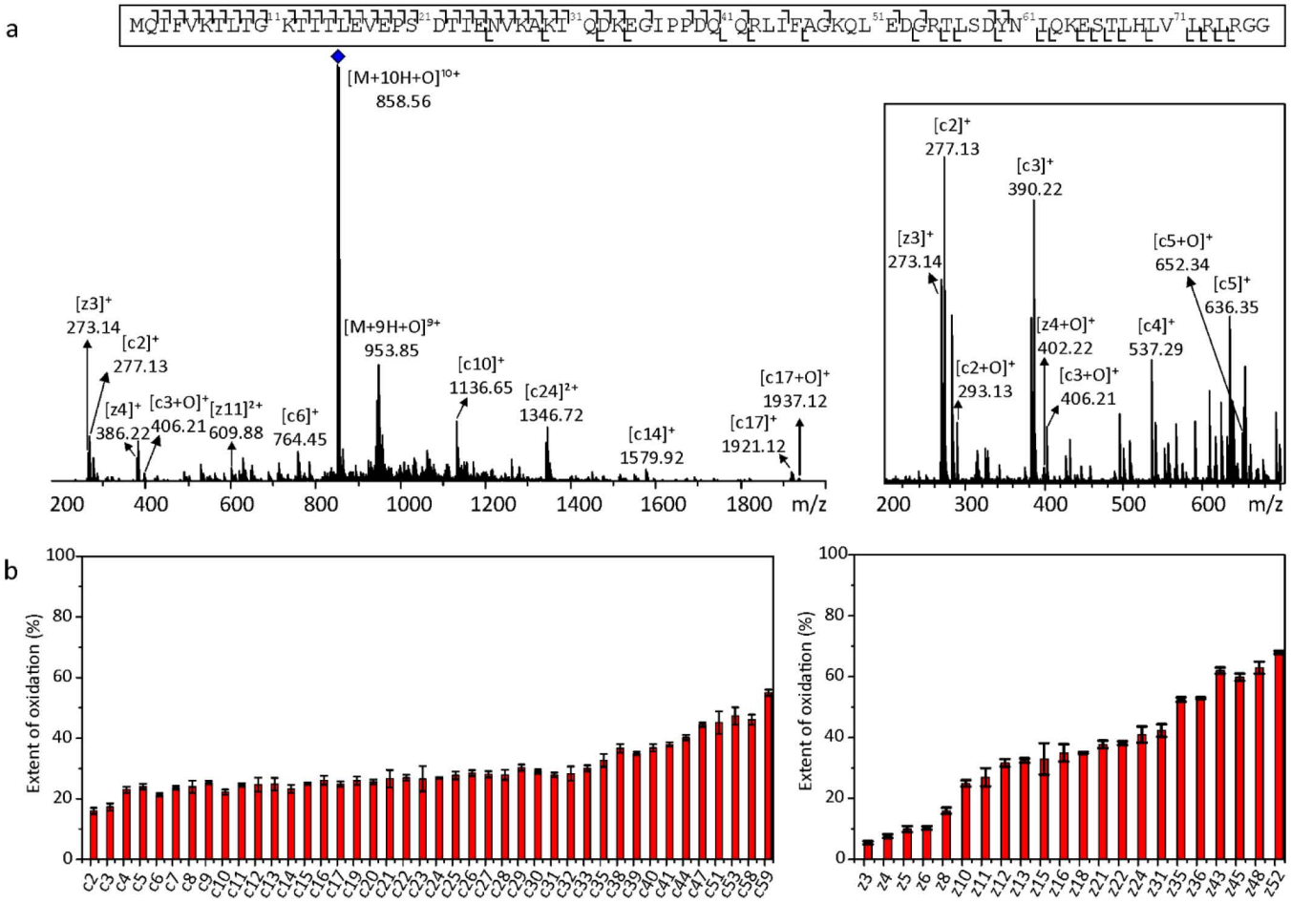


Figure 3: ECD spectrum of FPOP labeled ubiquitin. a) Overview of the whole spectrum obtained upon isolation of 10+ charge state singly oxidized ubiquitin (m/z 858.56) which is indicated with blue diamond, with zoom-in ECD spectrum of the m/z range 200-650. **b)** Extent of FPOP oxidation for ubiquitin on the fragment level. Red bars indicate oxidized fragment ions. Error bars are showing the standard deviations of three independent MS measurements.

The higher the charge state, the seemingly lower the ratio between oxidized and non-oxidized fragments and the calculated extent of oxidation.

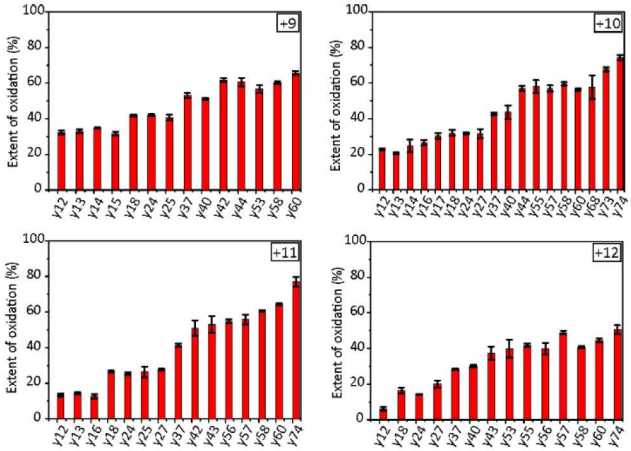


Figure 4: Comparison of the extent of oxidation for y ions at four charge states (9+, 10+, 11+ and 12+) of singly oxidized ubiquitin fragmented by CID. Red bars highlight the oxidized residues. Error bars are showing the standard deviations of three independent MS measurements.

Moreover, compared to ETD and ECD (**Figure S3 and S4**), the calculated extent of oxidation is lower in CID (see b2/c2 and y12/z12), which can be attributed to neutral loss of water from both b and y ions. (**Figure S5**). The loss of water was observed more frequently in oxidized fragments, which strongly affects the oxidation extent if calculated solely using ions without the water loss.

Multi CASI/CID and multi CASI/ECD were performed to overcome the variances in sequence coverage and oxidation extent obtained from differently charged ions. This technique allowed simultaneous isolation of several charge

states of the same ion in the quadrupole and transfer all these ions into ICR or collision cell for ECD or CID fragmentation, respectively. For both multi CASI/CID and multi CASI/ECD experiments, four different charge states (9+, 10+, 11+, and 12+) were selected and resulting fragmentation spectra and oxidation extents are shown in **Figure S6**. Both fragmentation techniques retain their characteristics in multi CASI mode, with CID having much lower sequence coverage and unstable oxidation levels due to loss of water and ECD with better sequence coverage and more stable oxidation extents.

As the next step, we tested the use of FPOP with a top-down approach to monitoring conformational changes in proteins. Well-studied horse heart myoglobin and its apoform (without heme group) served as a model protein. The holomyoglobin and apomyoglobin were separately exposed to hydroxyl radicals in FPOP experiment, and five different charge states (15+, 16+, 17+, 18+, and 19+) of singly oxidized protein were submitted to multi CASI/CID and multi CASI/ECD fragmentation (**Figure S7**). The resulting data were processed similarly to ubiquitin analysis, and the calculated extent of fragment ion oxidation is plotted in **Figure 5a** for CID and in **Figure 5b** for ECD. The combination of ions from CID and ECD provided sufficient sequence coverage for the determination of regions responsible for changes in oxidation, which reflect differences between myoglobin apo- and holoform structures. However, not all oxidation changes were significant. Therefore, we set the value of the minimum difference between two vicinal ions (2.5%) to avoid too sensitive impact, which could lead to false-positive results. The threshold corresponds to twice the average standard deviation of ion oxidation. Each fragment ion, which exceeds this minimal value, was highlighted in the myoglobin sequence (**Figure 5c**) to indicate a significant change in the structure. Red sequences represent more oxidized regions in apomyoglobin, while the greens show regions more oxidized in holomyoglobin. Further, overlapping sequence regions enable to localize oxidized residues more precisely, and the single amino acid resolution can be achieved, taking into account the amino acid side chains' susceptibility to hydroxyl radicals[25].

To facilitate a structural interpretation, the color-coded regions were mapped onto the X-ray structural model of holomyoglobin[26] (**Figure 5d and S8**). Our top-down data are in excellent agreement with previously published both apo- and holomyoglobin structural models. More oxidized apomyoglobin regions surround the heme cavity, and thus the cofactor release has the highest impact on the myoglobin structure. Moreover, our findings are also supported by previously published bottom-up FPOP experiment[27] where the observed differences in myoglobin structure are highlighted in violet (**Figure 5e**).

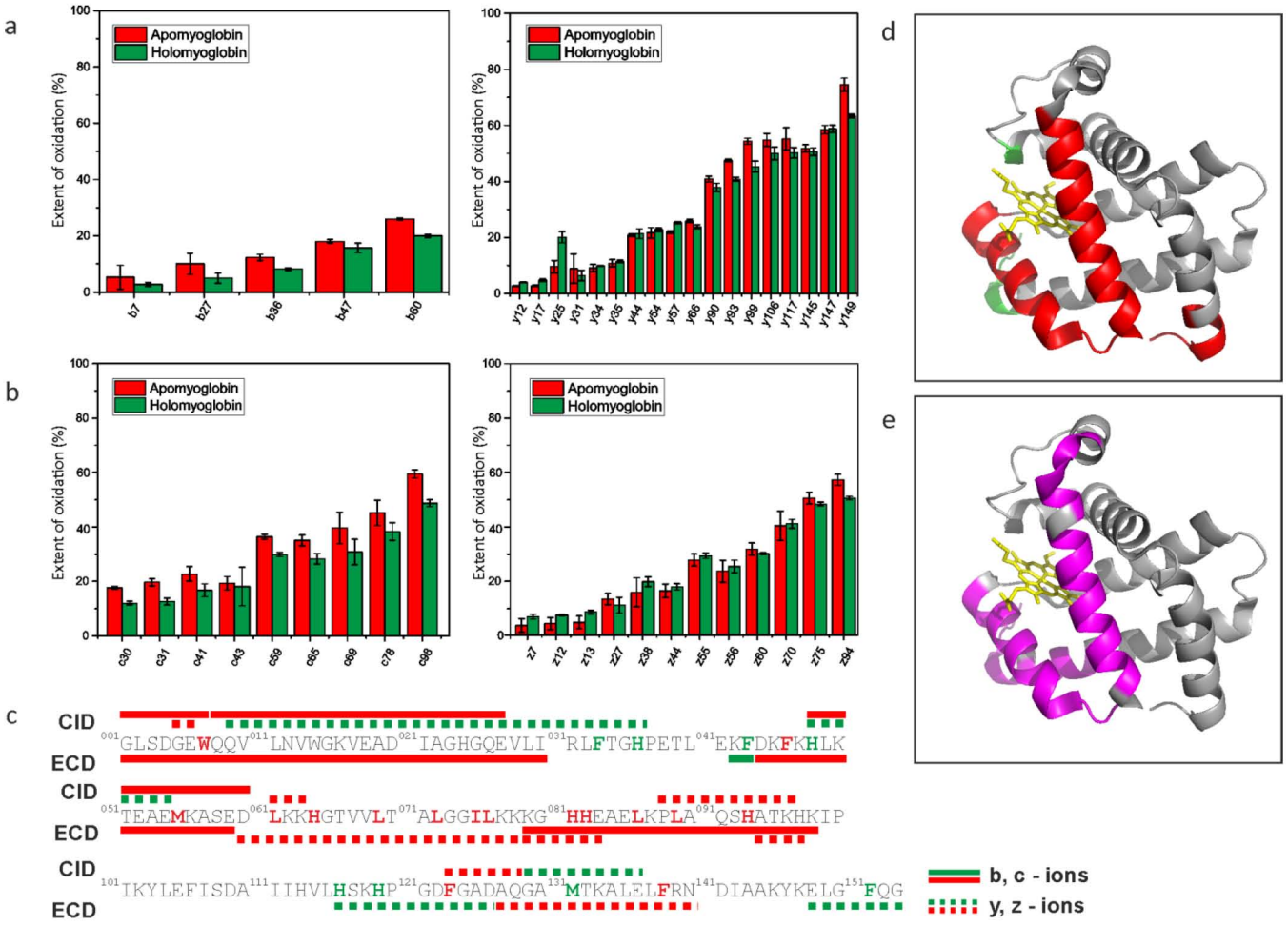


Figure 5. Extents of fragment ion oxidation resulting from multi CASI/CID and multi CASI/ECD MS/MS spectra obtained upon isolation of five charge states (15+, 16+, 17+, 18+ and 19+) of apomyoglobin and holomyoglobin. a) Extents of peptide oxidation calculated from multi CASI/CID and **b)** multi CASI/ECD. Red and green bars highlight the oxidized residues of apo- and holomyoglobin, respectively. Error bars are showing the standard deviations of three independent MS measurements. **c)** Myoglobin sequence with highlighted peptides and residues from multi CASI/CID and multi CASI/ECD whose oxidation is significantly increased in apomyoglobin (in red) or holomyoglobin (in green). **d)** Crystal structures of native holomyoglobin (1WLA)[26] with visualized regions most affected by heme removal as detected by our FPOP top-down approach or **e)** by classical FPOP bottom-up approach[27] where the biggest differences in oxidation are highlighted in violet.

CONCLUSION

We show that top-down mass spectrometry can be used in FPOP to obtain detailed information from a specific population of labeled protein. In our case, it is singly oxidized ubiquitin by isolating and fragmenting it using top-down compatible fragmentation techniques. CID, ETD, and ECD fragmentation techniques were tested, and results were analyzed with our home-built software. Unfortunately, CID and ETD fragmentations suffer from low fragmentation yield resulting in poor sequence coverage, in addition to oxidation extents measured from CID being affected by the neutral loss of water in CID, preferentially occurring in oxidized fragment ions and thus biasing the data. The use of ECD improved the sequence coverage, which is extremely important to pinpoint the site of oxidation to a specific residue. Since the oxidation extent from FPOP labeling was somewhat dependent on the charge state of precursor ion, multi CASI/CID and multi CASI/ECD fragmentation techniques were used to isolate and fragment multiple charge states of oxidized protein to alleviate this bias partially. Moreover, we proved that the combination of FPOP and top-down mass spectrometry could also be used for monitoring of conformational changes in proteins, and multi CASI/ECD was found to be the most suitable of all tested fragmentation techniques. This approach can therefore be

employed in the future to reliably elucidate structural changes in proteins under different conditions by comparing the FPOP-induced labeling changes.

METHODS

Material and reagents: hydrogen peroxide (H_2O_2) (30%), methionine, ubiquitin, holomyoglobin, and ammonium bicarbonate were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO) in the highest available purity. Apomyoglobin was prepared from holomyoglobin using previously published butanone extraction[28].

Fast photochemical oxidation of proteins: FPOP was performed as described in the previous publications[29,30]. The FPOP setup consists of three syringes loaded into two syringe pumps (New Era Pump System Inc., Model NE-1000, and Model NE-4000), creating a continuous flow. The first syringe was filled with protein (ubiquitin or myoglobin) in 50 mM degassed ammonium bicarbonate buffer (pH 7.2), the second syringe contained H_2O_2 (final concentration of 0.03%), and the third syringe contained 150 mM methionine as a quencher of the reaction. H_2O_2 was mixed with the protein in the first segment of the flow apparatus, the photolysis of H_2O_2 was then performed by a 248 nm KrF excimer laser (Coherent, COMPex50) with the energy of 50 mJ per pulse and frequency 50 Hz (100 mJ and 10 Hz in case of myoglobin). The laser was irradiating the sample in a 75-μm i.d. fused silica tubing (Polymicro Technologies) through a transparent window prepared by removing approximately 3 mm segment of the polyimide coating of the tubing. After H_2O_2 photolysis, the sample was mixed with methionine from the third syringe to stop the labeling reaction. Syringe pumps for sample and H_2O_2 were set at 10 μL min^{-1}, while methionine syringe flow rate was 20 μL min^{-1}. Sample was collected for 1 min.

Mass spectrometry: Samples were desalted on protein micro trap column (OPTI-TRAP micro, Optimize technologies) using 0.1% formic acid (three times by 250 μl) and eluted (50 μl of 80% acetonitrile, 0.1% formic acid; one repeat). All mass spectra were obtained on Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer (15 T solariX XR, Bruker Daltonics) equipped with a home-built nanospray source operated in the positive mode with a desolvation temperature of 200 °C. The nanospray source consists of quartz capillaries (1.0 mm O.D./0.70 mm I.D.) pulled with a CO_2 laser-based micropipette puller (Model P-2000, Sutter Instrument Co., Novato, CA). The nanoelectrospray (nESI) was performed by applying ~ 400–600 kV voltage to the platinum wire inserted into the capillary with the sample solution (~ 5 μL samples were loaded into the capillary using a gel-loader pipette tip). The voltage was optimized for each individual sample to compensate for differences in capillary length and distance from the mass spectrometer inlet. After the desolvation step, the ions were transferred to the quadrupole where singly oxidized protein (charge states 9+, 10+, 11+, and 12+ for ubiquitin or 15+, 16+, 17+, 18+, and 19+ for myoglobin) was isolated with an isolation window of 1.5 Da and then transferred to a hexapole (CID and ETD) or ICR cell (ECD) for fragmentation. Ion accumulation was set to 1 s for CID and 0,05 s for ETD and ECD[31,21,22]. The time of flight was set at 0.9 ms, and the m/z range used was 150–2500 m/z at 2 million data points. For all MS/MS analyses, 128 scans in triplicate were acquired. For CID experiments, both the collision energy and the accumulation time in the hexapole collision cell were optimized for each precursor ion. The collision energy used was between 8 to 14 eV, depending on the ion charge state. For ECD, electron pulse length was set to 0.08 s, filament bias voltage to 0.7 V, ECD lens to 12 V, and cathode heater current was set to 1.5 A. The ETD experiments were performed using a nCI source, which was tuned to maximize the output of ETD reagent ion (m/z 202) using ionization energy of 70 eV and emission current of 3.5 μA. The reagent accumulation was set to 150 ms and reaction time to 100 ms.

For the multi CASI experiment, a full MS scan was first performed, and then singly oxidized protein ions were isolated in charge stats from 9+ to 12+ for ubiquitin or from 15+ to 19+ for myoglobin with individual isolation windows. In the case of multi CASI/CID, ions were accumulated for 1 s in the collision cell, and different collision voltages were set for each individual isolation window to induce fragmentation. Mass spectra were averaged over 128 scans. In the case of multi CASI/ECD, multiple charge states are selected and transferred to the ICR cell and the ions were irradiated with electrons for 0.08 s with a filament bias voltage 0.7 V and cathode heater current 1.5 A.

Data analysis: Oxidation level quantitation was calculated by the fragment ion intensities from ETD, ECD, and CID to determine the oxidation extent of a given fragment. DataAnalysis 5.0 software (Bruker Daltonics) was used for the visualization and calibration of ESI-FTICR spectra. Theoretical m/z values of CID, ETD, ECD, multi CASI/CID and multi CASI/ECD fragment ions were obtained using the Ms2links tool. Firstly, oxidation yields for all the ions

were obtained manually as the ratio of the oxidized fragment form to the sum of both oxidized and non-oxidized forms (eq.1)[32].

$$\text{eq.1} \qquad Extent\ of\ oxidation = \frac{\sum I_{ox}}{\sum (I_{ox} + I)}$$

I_{ox} denotes the intensity of the oxidized and I the intensity of the non-oxidized fragment ion. Average oxidation level quantitation from multiple charge states of oxidized protein in multi CASI was calculated as a ratio of the sum of oxidized fragment intensity in all of the charge states to the sum of the intensity of oxidized and non-oxidized fragment in all of the charge states.

$$\text{eq.2} \qquad p = \frac{\sum I_{ox}(n) + \sum I_{ox}(n+1) + \cdots}{\sum (I_{ox} + I\,(n)) + \sum (I_{ox} + I\,(n+1)) + \cdots}$$

P denotes the average oxidation, I_{ox} denotes the intensity of the oxidized fragment ion, and I refers to the intensity of *the* non-oxidized fragment ion. The n corresponds to the individual charge states.

ASSOCIATED CONTENT

Supporting Information

The supporting information contains mass spectra, bar plots for the extent of oxidation from CID, ECD, and ETD, zoomed-in CID mass spectra for b18 ion, multi CASI CID and ECD spectra for oxidized ubiquitin and holomyoglobin and structure of holomyoglobin with oxidized residues labeled by FPOP. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

*Petr Novak – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0001-8688-529X;E-mail: pnovak@biomed.cas.cz. Fax: +420 241 062 156. Tel.: +420 325 873 610 (P.N.).

Authors

Ghazaleh Yassaghi – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0002-0422-0671

Zdenek Kukacka – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0001-7569-843X

Jan Fiala – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0002-7753-8170

Daniel Kavan – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0003-2693-1199

Petr Halada – Institute ofMicrobiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; orcid.org/0000-0002-7229-3450

Author Contributions

The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

§These authors contributed equally. (match statement to author names with a symbol)

ACKNOWLEDGMENT

We thank Frantisek Filandr for his help during the writing of manuscript. This work was mainly financially supported by the Czech Science Foundation (19-16084S) and the European Commission H2020 (EU_FT-ICR_MS grant agreement ID: 731077 and EPIC-XS - grant agreement ID: 823839). Additional institutional and facility support from the Academy of Sciences of the Czech Republic (RVO: 61388971), the Ministry of Education of the Czech Republic (structural mass spectrometry CF - LM2018127 CIISB), and the European Regional Development Funds (CZ.1.05/1.1.00/ 02.0109 BIOCEV) are gratefully acknowledged.

REFERENCES

(1) Wang, L.; Chance, M. R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol. Cell. Proteomics* **2017**, *16* (5), 706–716.

(2) Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, *120* (10), 4355–4454.

(3) Maleknia, S. D.; Ralston, C. Y.; Brenowitz, M. D.; Downard, K. M.; Chance, M. R. Determination of Macromolecular Folding and Structure by Synchrotron X-Ray Radiolysis Techniques. *Anal. Biochem.* **2001**, *289* (2), 103–115.

(4) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2057–2063.

(5) Zhang, B.; Rempel, D. L.; Gross, M. L. Protein Footprinting by Carbenes on a Fast Photochemical Oxidation of Proteins (FPOP) Platform. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (3), 552–555.

(6) Chance, M. R.; Sclavi, B.; Woodson, S. A.; Brenowitz, M. Examining the Conformational Dynamics of Macromolecule with Time-Resolved Synchrotron X-Ray "Footprinting." *Structure* **1997**, *5* (7), 865–869.

(7) Heyduk, E.; Heyduk, T. Mapping Protein Domains

Involved in Macromolecular Interactions: A Novel Protein Footprinting Approach. [Erratum to Document Cited in CA121:103552]. *Biochemistry* **2005**, *34* (46), 15388–15388.

(8) Aye, T. T.; Low, T. Y.; Sze, S. K. Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* **2005**, *77* (18), 5814–5822.

(9) Gau, B. C.; Chen, H.; Zhang, Y.; Gross, M. L. Sulfate Radical Anion as a New Reagent for Fast Photochemical Oxidation of Proteins. *Anal. Chem.* **2010**, *82* (18), 7821–7827.

(10) Zhang, M. M.; Rempel, D. L.; Gross, M. L. A Fast Photochemical Oxidation of Proteins (FPOP) Platform for Free-Radical Reactions: The Carbonate Radical Anion with Peptides and Proteins. *Free Radic. Biol. Med.* **2019**, *131*, 126–132.

(11) Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chemie - Int. Ed.* **2017**, *56* (45), 14007–14010.

(12) Stadtman, E. R.; Levine, R. L. Free Radical-Mediated Oxidation of Free Amino Acids and Amino Acid Residues in Proteins. *Amino Acids* **2003**, *25* (3–4), 207–218.

(13) Xu, G.; Chance, M. R. Radiolytic Modification and Reactivity of Amino Acid Residues Serving as Structural Probes for Protein Footprinting. *Anal. Chem.* **2005**, *77* (14), 4549–4555.

(14) Limpikirati, P.; Liu, T.; Vachet, R. W. Covalent Labeling-Mass Spectrometry with Non-Specific Reagents for Studying Protein Structure and Interactions. *Methods* **2018**, *144*, 79–93.

(15) Liu, X. R.; Zhang, M. M.; Zhang, B.; Rempel, D. L.; Gross, M. L. Hydroxyl-Radical Reaction Pathways for the Fast Photochemical Oxidation of Proteins Platform As Revealed by 18 O Isotopic Labeling . *Anal. Chem.* **2019**.

(16) Rozbeský, D.; Rosůlek, M.; Kukačka, Z.; Chmelík, J.; Man, P.; Novák, P. Impact of Chemical Cross-Linking on Protein Structure and Function. *Anal. Chem.* **2018**, *90* (2), 1104–1113.

(17) Pamreddy, A.; Panyala, N. R. Top-down Proteomics : Applications , Recent Developments and Perspectives. **2016**, *2* (2), 52–75.

(18) Kruppa, G. H.; Schoeniger, J.; Young, M. M. A Top down Approach to Protein Structural Studies Using Chemical Cross-Linking and Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (2), 155–162.

(19) Novak, P.; Young, M. M.; Schoeniger, J. S.; Kruppa, G. H. A Top-down Approach to Protein Structure Studies Using Chemical Cross-Linking and Fourier Transform Mass Spectrometry. *Eur. J. Mass Spectrom.* **2003**, *9* (6), 623–631.

(20) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. A Top-down Method for the Determination of Residue-Specific Solvent Accessibility in Proteins. *J. Mass Spectrom.* **2004**, *39* (3), 322–328.

(21) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2004**, *101* (26), 9528–9533.

(22) Zubarev, R.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc* **1998**, *120* (16), 3265–3266.

(23) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. High Throughput Protein Fold Identification by Using Experimental Constraints Derived from Intramolecular Cross-Links and Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (11), 5802–5806.

(24) Cooper, H. J.; Håkansson, K.; Marshall, A. G. The Role of Electron Capture Dissociation in Biomolecular Analysis. *Mass Spectrom. Rev.* **2005**, *24* (2), 201–222.

(25) Xu, G.; Chance, M. R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007**, *107* (8), 3514–3543.

(26) Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; Mauk, A. G.; Smith, M.; Brayer, G. D. A Myoglobin Variant with a Polar Substitution in a Conserved Hydrophobic Cluster in the Heme Binding Pocket. *Biochim. Biophys. Acta* **1997**, *1341*, 1–13.

(27) Vahidi, S.; Stocks, B. B.; Liaghati-Mobarhan, Y.; Konermann, L. Mapping PH-Induced Protein Structural Changes under Equilibrium Conditions by Pulsed Oxidative Labeling and Mass Spectrometry. *Anal. Chem.* **2012**, *84* (21), 9124–9130.

(28) Teale, F. W. Cleavage of the Haem-Protein Link by Acid Methylethylketone. *Biochim. Biophys. Acta* **1959**, *35* (20), 543.

(29) Zhang, Y.; Rempel, D. L.; Zhang, H.; Gross, M. L. An Improved Fast Photochemical Oxidation of Proteins (FPOP) Platform for Protein Therapeutics. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (3), 526–529.

(30) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**.

(31) McLafferty, F. W.; Bente, P. F.; Kornfeld, R.; Tsai, S. ???C; Howe, I. Collisional Activation Spectra of Organic Ions. *J. Mass Spectrom.* **1995**, *30* (6), 797–806.

(32) Hambly, D.; Gross, M. Laser Flash Photochemical Oxidation to Locate Heme Binding and Conformational Changes in Myoglobin. *Int. J. Mass Spectrom.* **2007**, *259* (1–3), 124–129.

ARTICLE II

Loginov D., Fiala J., Chmelík J., Brechlin P., Kruppa G., Novák P.

Benefits of Ion Mobility Separation and Parallel Accumulation–Serial Fragmentation Technology on timsTOF Pro for the Needs of Fast Photochemical Oxidation of Protein Analysis.

ACS Omega 2021, 6, 15, 10352–10361

My contribution: Conducting experiments, LC-MS/MS analysis, writing methodological part of manuscript and creating figures

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

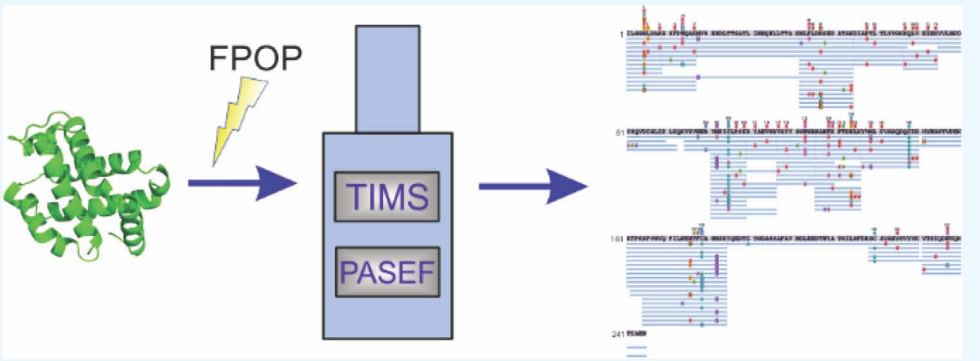
Benefits of Ion Mobility Separation and Parallel Accumulation–Serial Fragmentation Technology on timsTOF Pro for the Needs of Fast Photochemical Oxidation of Protein Analysis

Dmitry S. Loginov,* Jan Fiala, Josef Chmelik, Peter Brechlin, Gary Kruppa, and Petr Novak*

Cite This: *ACS Omega* 2021, 6, 10352–10361

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Fast photochemical oxidation of proteins (FPOP) is a recently developed technique for studying protein folding, conformations, interactions, etc. In this method, hydroxyl radicals, usually generated by KrF laser photolysis of H_2O_2, are used for irreversible labeling of solvent-exposed side chains of amino acids. Mapping of the oxidized residues to the protein's structure requires pinpointing of modifications using a bottom-up proteomic approach. In this work, a quadrupole time-of-flight (QTOF) mass spectrometer coupled with trapped ion mobility spectrometry (timsTOF Pro) was used for identification of oxidative modifications in a model protein. Multiple modifications on the same residues, including six modifications of histidine, were successfully resolved. Moreover, parallel accumulation–serial fragmentation (PASEF) technology allows successful sequencing of even minor populations of modified peptides. The data obtained indicate a clear improvement of the quality of the FPOP analysis from the viewpoint of the number of identified peptides bearing oxidative modifications and their precise localization. Data are available via ProteomeXchange with identifier PXD020509.

INTRODUCTION

Methods of structural mass spectrometry, such as hydrogen deuterium exchange (HDX-MS), chemical cross-linking (XL-MS), and various chemical labeling methods, have become a very valuable and effective tool for the study of protein structures and protein–ligand and protein–protein interactions. Among them, fast photochemical oxidation of protein (FPOP) employing hydroxyl radicals, introduced by Aye and co-workers[1] and later by Hambly and Gross,[2] has been recognized for its ability to study conformational changes in proteins.[3] The main advantages of this method are the speed of the labeling reaction, which occurs on the time scale of microseconds, and its broad reactivity.[4] FPOP irreversibly labels solvent-accessible amino acid side chains and allows the determination of reactivity at single residue resolution more easily in comparison with hydrogen deuterium exchange.[5]

In theory, side chains of all amino acids except Gly could be modified by ·OH.[6] However, stable oxidation products have been detected by mass spectrometry for only 14 of them.[4] Recently, mechanisms of oxidative modifications of 13 amino acids have been reported.[7] The amino acids have been classified into three categories with respect to mechanisms of oxidation based on the source of oxygen uptake. Class 1 consists of His, Arg, Tyr, and Phe, which preferably take oxygen from H_2O_2. Met is assigned to class 2, where oxygen is competitively taken from H_2O_2 and dissolved oxygen. The rest of the reactive amino acids (Leu, Ile, Val, Pro, Lys, Asp, Gln, and Glu) comprise class 3 and take oxygen only from O_2.[7]

The most reactive amino acids using FPOP are sulfur-containing Cys and Met followed by the aromatic ones with the order of reactivity Trp > Tyr > Phe.[3] It should be noted

Received: February 9, 2021
Accepted: March 29, 2021
Published: April 8, 2021

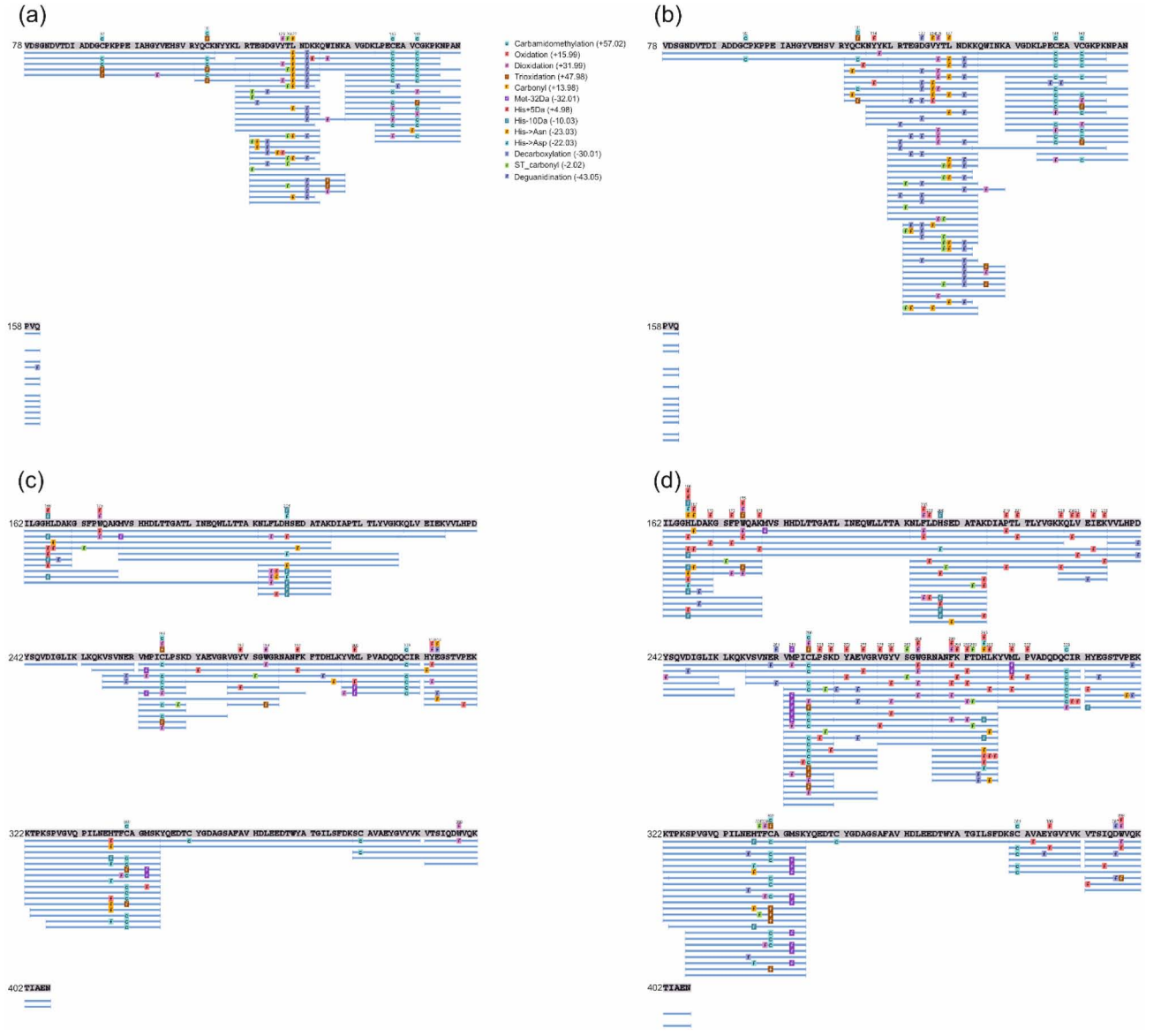


Figure 1. Peptide map of haptoglobin (40 ng): Hpα sequence coverage for (a) laser off and (b) laser on samples. Hpβ sequence coverage for (c) laser off and (d) laser on samples. Peptides passing thresholds −10lgP > 15 and AScore > 20 were accepted for analysis. Modifications localized on a particular residue from fragmentation spectra are indicated above the sequence line.

that some amino acids have several possible modifications, which should be considered for a more accurate analysis.[6] Sometimes, to simplify data analysis, only major modifications, like oxidation to carbonyl (+14 Da), oxidation (+16 Da), and di- and trioxidation, might be taken into account,[8] but some valuable information may be lost in this approach.

Precise localization of a modified residue is of great importance in FPOP analysis, representing a very challenging task. Because of the broad reactivity of hydroxyl, a single peptide might have the same modification at different residues. Another obstacle arises from the dynamic range in intensities of an unmodified peptide and its oxidized form, which can be up to 3 orders of magnitude.[9] Thus, instrumentation plays a crucial role in FPOP analysis.

From this point of view, the timsTOF Pro mass spectrometer from Bruker Daltonics has performance characteristics that should be very useful. The timsTOF Pro is a quadrupole time-of-flight instrument with integrated trapped ion mobility spectrometry (TIMS) allowing a third dimension of peptide separation.[10] Another advantage of this device is the "parallel accumulation−serial fragmentation" (PASEF) method.[11] The combination of both TIMS separation and PASEF allows selection of multiple precursors with the same mass and retention time but differing mobilities for fragmentation and increases the sequencing speed without loss of sensitivity.[12] Good performance of this device has been already shown in the shotgun proteomic approaches.[12,13]

In the current study, we present data indicating that the timsTOF Pro mass spectrometer has great potential to improve FPOP analyses.

RESULTS AND DISCUSSION

Before the FPOP analysis, the purity of Hp was confirmed by SDS-PAGE electrophoresis. Only two bands corresponding to the α and β subunits were resolved on the gel (Supporting Information, Figure S1), allowing the use of a simplified database for searching MSMS data.

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

Two injection volumes, corresponding to 20 and 40 ng of digested protein, were analyzed. Using a bottom-up approach, a nearly 100% sequence coverage of the model protein Hp was achieved with 20 ng of the peptide mixture (Supporting Information, Figure S2). A twice higher injection volume led to full sequence coverage of the protein and a significant increase in the number of modified peptides identified.

A comparison of peptide maps of control (laser off) and irradiated (laser on) samples revealed a higher number of modifications in the latter case for both Hp subunits (Figure 1). Four and 16 modifications on 4 and 7 types of amino acids were unambiguously deduced from fragmentation spectra in Hpα and Hpβ, respectively, without laser exposure, whereas 6 and 55 such modifications on 5 and 15 types of amino acids were found in the respective samples after laser irradiation. Hpβ was modified at a much higher extent than Hpα, making it more informative for the evaluation of the timsTOF Pro for FPOP analysis (Table 1 and Supporting Information, Table S2).

Table 1. Amino Acids Modified with Hydroxyl Radicals during the FPOP Experiment

	position	
residue	Hpα	Hpβ
Cys	111	266, 340
Trp		175, 284, 398
Tyr	114, 125	272, 280, 386
Met		263, 300
Phe		173, 205, 290, 292, 339
His		166, 208, 295
Arg		261
Ile		233
Leu	127	167, 206, 221, 230, 296
Val	124	231, 275, 278
Pro		219, 268, 302
Thr		293, 338
Ser		282
Lys		170, 178, 228, 235, 270, 291
Asp	122	397

Plotting of all certain modifications detected in the irradiated sample onto the Hp 1-1 3D (PDB 4WJG) structure revealed that oxidized residues were in close proximity to the protein surface (Figure 2). This observation indirectly indicates the reliability of identification of modified residues with the selected approach. Also, a correlation between the solvent-accessible surface area (SASA) and modification rates of amino acids was evaluated (Figure 3). As it was expected, the increase of SASA led to a higher modification extent of more reactive residues, like aromatic amino acids and Met. It should be mentioned that an Hp 1-1 high-resolution structural model has not been deposited in its free form so far. Thus, Hp coordinates for SASA calculation were derived from the Hb−Hp complex stabilized with the *Trypanosoma brucei brucei* receptor (PDB 4WJG).[14] Therefore, detected oxidation of Trp175 and Tyr272 with calculated zero SASA and minor oxidation of highly accessible His295 might indicate different positions of the respective side chains of Hp 1-1 in its free form. Also, these data are in agreement with the reactivity of amino acids in the FPOP analysis.[6] Slightly accessible, highly reactive Tyr, Trp, Phe, or Met residues were readily oxidized by ·OH, whereas low- or middle-reactive amino acids, namely,

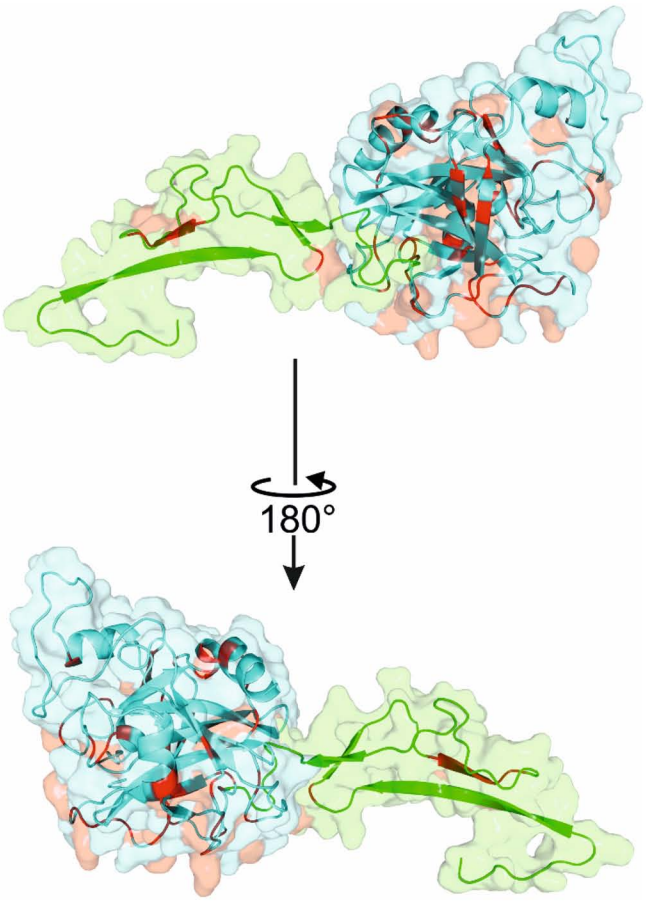


Figure 2. Structural model of Hp 1-1 (based on PDB 4WJG) with highlighted modified residues identified in the laser irradiated sample. The α subunit is in green color, and the β subunit is in cyan color. Modified residues are in red color. Respective light colors represent the solvent-accessible area of the protein surface.

Leu, Ile, Val, and Lys, were less modified, even possessing high SASA values (Figure 3).

Analysis of the identified modifications showed oxidation (+16 Da) to be the dominant one. Also, we were able to easily assign rare modifications on Thr residues that were previously shown to be difficult to detect.[3] Thr modification was assigned by PEAKS+ software to the peptide with the sequence NANFKFTDHLK (Supporting Information, Figure S3). The manual validation of the raw spectrum confirmed PEAKS+ assignment (Figure 4). An unmodified N-terminal sequence was deduced from the respective b-ion series b_3−b_6. Fragment b_6 at m/z 722.35 corresponded to the unmodified NANFKF sequence. The loss of 2 a.m.u. was detected on the b_7 ion at m/z 821.38, which corresponded to the NANFKFT sequence, meaning carbonylation on the Thr residue. The subsequent b_8 at m/z 936.42 and b_9 at 1056.55 fragment ions confirmed such observation. Thr oxidative modification was also supported by the complementary y-ion series, where y_4 at m/z 512.28 represented the unmodified DHLK sequence and all further detected y fragments (y_6−y_9) were bearing −2.02 a.m.u. modification (Figure 4).

For some amino acids, several types of modifications were detected, generating different populations of specific peptides. For instance, mono-, di-, or trioxidized forms of aromatic amino acids (Trp, Tyr, or Phe) were found as well as di- and trioxidized Cys (Figure 1). Also, five possible histidine

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

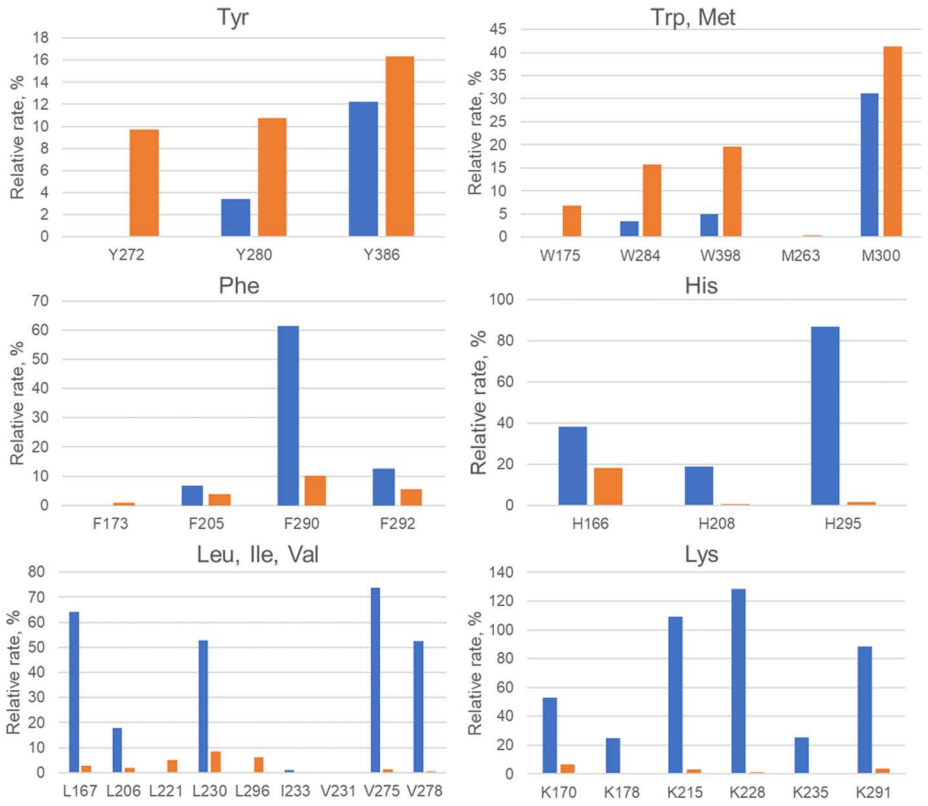


Figure 3. Correlation of oxidation extents (orange bars) of single amino acids and their relative solvent-accessible surface areas (blue bars) calculated using the Hp 1-1 coordinates extracted from the PDB 4WJG model.

modifications that may occur during FPOP experiments described in previous work were detected in the present study (Table 2).[6] Using PASEF technology, it was possible to get direct evidence of differently modified His residues from the fragmentation spectra (Figure 5) even for very low intensity precursors (Table 2). Moreover, one more possible modification of His, namely, +14 Da,[15] was determined (Figure 5f).

The most abundant histidine modification detected in both samples was +16 Da followed by +14 and −10 Da (Table 2). However, a contribution of peptides with −10 and +14 Da modifications to the overall oxidation of the peptide was higher for the sample without laser irradiation, indicating that these oxidations are from H_2O_2. The remaining three modifications (His to Asn or Asp, +5 Da) represented minor populations and were characteristic of oxidation with ·OH.

The introduction of trapped ion mobility spectrometry on the timsTOF Pro instrument provides an additional dimension of peptide separation. Several products of hydroxyl radical footprinting were detected at the same retention time, and the ion mobility separation resolved them, which is crucial for the quantification of modified products (Supporting Information, Table S3). The representative example is demonstrated in Figure 6. Using the ion mobility separation, two forms of the peptide with the sequence NLFLDHSEDATAK in its doubly charged form at m/z 738.84 and the same RT (9.44 min) were identified. The more abundant form showed oxidation of the Phe residue, and the minor form was assigned to the oxidation of the His residue. Both modified peptides were identified by PEAKS+ software with a high score: 113.76 and 84.83 for oxidized Phe and His, respectively. Oxidation of Phe was deduced from the respective b and y fragment ion series. The mass difference of 163.06 a.m.u. between b_2 at m/z 228.13 and b_3 at m/z 391.19 together with y_{11} at m/z 1249.55 and y_{10} at m/z 1086.49 fits exactly the mass of the Phe residue (147.07 a.m.u.) plus oxygen (+15.99 a.m.u.). In the case of His modification, b_5 at m/z 603.29 corresponded to the unmodified NLFLD N-terminal peptide sequence while b_7 at m/z 972.42 reflected oxidation of the peptide region containing His and Ser. The complementary y_6° at m/z 616.28 and at y_8° at m/z 856.36 confirmed the oxidation of the same sequence. The oxidation was finally assigned to His since the oxidation of Ser is extremely rare, and the rate constant of His oxidation is an order of magnitude higher than that of Ser.[6]

Previously, sub-residue resolution in an FPOP experiment has been reported.[16] Analogue results were obtained in the present study, where peptides containing aromatic amino acids oxidized at different sites were found. In particular, fragmentation spectra of peptides with position numbers of amino acids 203−215 (m/z 738.8416) eluting at 7.7 and 9.4 min were identical and corresponded to the oxidized Phe (Supporting Information, Figure S4). Similar fragmentation spectra were also found for the peptides 271−277 (m/z 825.3568, RT 5.6 and 6.5) and 392−401 (m/z 610.3093, RT 9.7, 10.4, and 12.4) bearing oxidative modification of Tyr and Trp, respectively (Supporting Information, Figures S5 and S6). Nevertheless, the key factor in achieving sub-residue resolution is the quality of liquid chromatography (LC) separation.

Five possible sites of oxidation were found in the peptide with sequence KQLVEIEK at m/z 501.78. Modification of different residues leads to changes in retention times of the peptides (Table 3). Another indication supporting oxidation of multiple sites in the peptide was changes in the ion mobility parameter. Another benefit of the PASEF technology is that

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

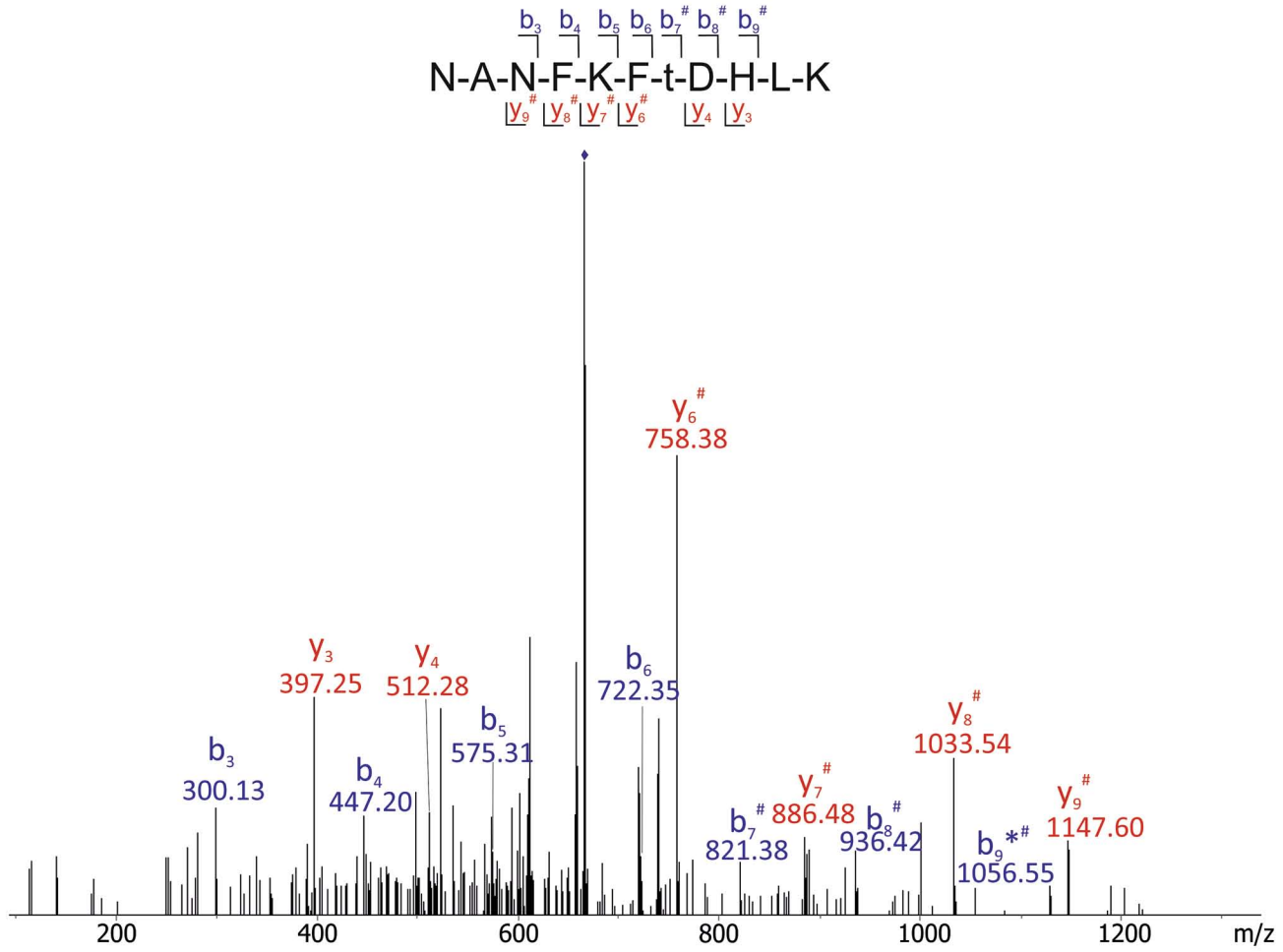


Figure 4. Fragmentation spectrum of the peptide with the sequence NANFKFTDHLK (doubly charged at m/z 666.8388) with a modification of a Thr residue identified during FPOP analysis. A precursor ion is labeled with a blue rhombus. Pound symbol, oxidized fragment ion. The modified residue is labeled with a lowercase letter.

Table 2. List of Detected Histidine Modifications in the ILGGHLDAK Peptide[a]

m/z (no laser/laser)	−10lgP (no laser/laser)	Z (no laser/laser)	RT, min (no laser/laser)	$1/k_0$ range (no laser/laser)	intensity* (no laser/laser)	contribution to the overall oxidation of the peptide* (no laser/laser), %	modification
469.24	34.28/32.24	2	9.45	0.8353−0.8539	13,852/22,520	1.2/0.6	+14 Da
900.49	53.31	1	8.59	1.4441−1.4619	-/41,104	-/1.2	His → Asn
901.48	51.15	1	9.47	1.4377−1.4555	-/27,527	-/0.8	His → Asp
913.48	49.49/43.38	1	6.21	1.4477−1.4655	8763/11,365	0.7/0.3	−10 Da
928.48	52.13	1	9.23	1.4527−1.4705	-/21,471	-/0.6	+5 Da
470.25/939.50	35.18/62.14	2/1	9.06/9.01	0.8182−0.8368 /1.4833−1.5010	6025/505,579	0.5/14.5	+16 Da

[a] $1/k_0$, reduced ion mobility coefficient;[12] *, the parameter was calculated using all charge states of respective peptides.

the fragmentation spectra for peptides can be obtained over a wide dynamic range of respective intensities, which has been shown previously in the proteomic data.[12] In the present study, this range was estimated to be almost 3 orders of magnitude (Table 3). Moreover, the quality of fragmentation spectra allowed a precise determination of modified amino acid even in the least abundant peptide (Figure 7).

Quantification is a key point of FPOP analysis for the precise footprinting of a protein surface. Thus, its reproducibility is of crucial importance. To evaluate this parameter, we selected the KQLVEIEK peptide bearing various modifications with different intensities representing high-, middle-, and low-intensive populations (Figure 8).

Obtained data confirmed a high accuracy of the timsTOF Pro even for low abundant peptides. It should be mentioned that these results were highly reproducible, including retention time, m/z, and ion mobility ($1/k_0$), meaning a confident automatic data interpretation in the case of multiplicates (Supporting Information, Figure S7).

■ CONCLUSIONS

In the present study, a benefit of ion mobility application for the needs of FPOP analysis was demonstrated for the first

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

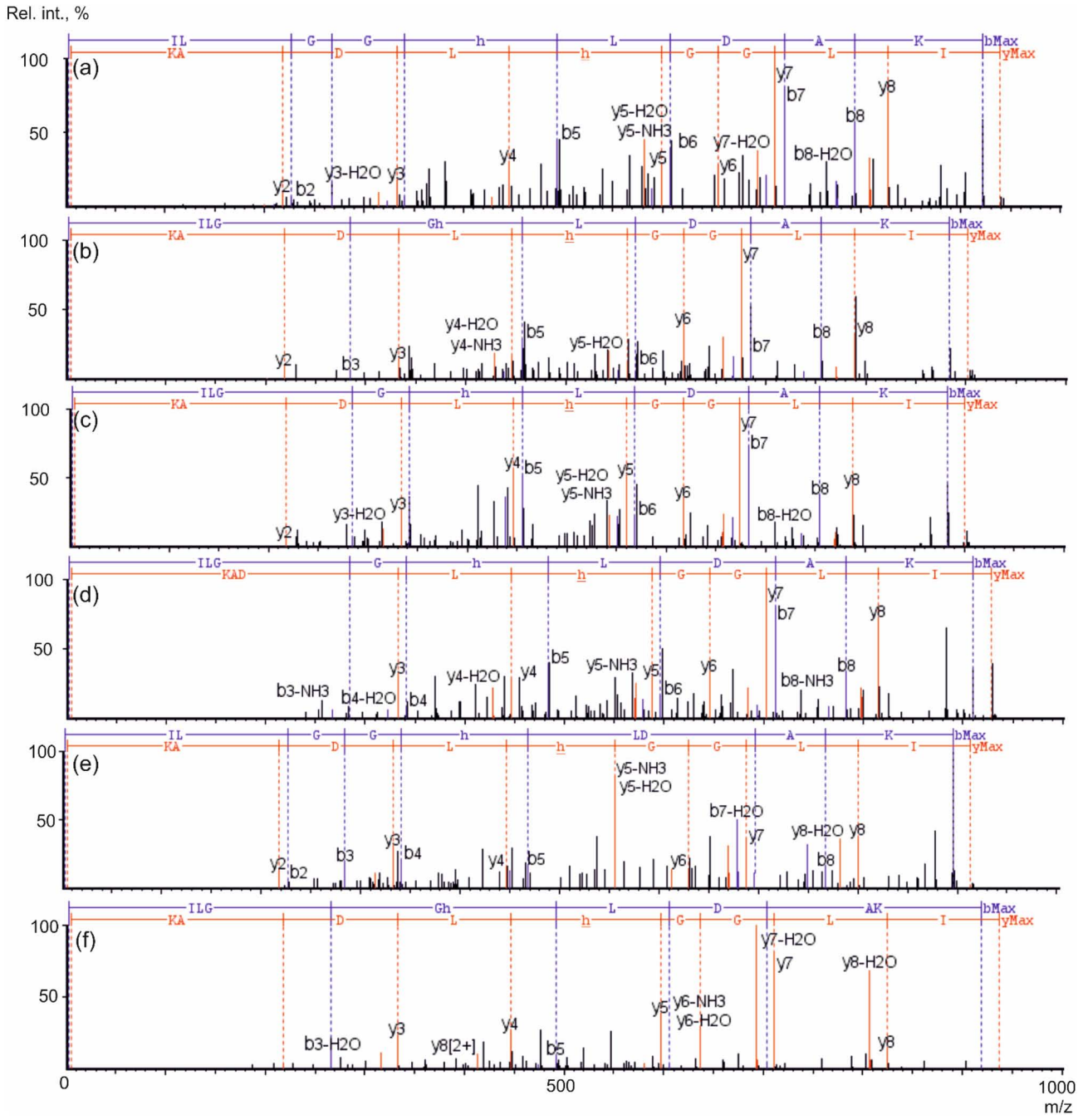


Figure 5. Fragmentation spectra of the peptide with the sequence ILGGHLDAK with different modifications at the histidine residue: (a) +16 Da, (b) His → Asp, (c) His → Asn, (d) +5 Da, (e) −10 Da, and (f) +14 Da. Modified residues are labeled with an underlined lowercase letter.

time. We demonstrate that TIMS and PASEF technologies in the timsTOF Pro instrument produce reliable data. Multiple modifications on the same residues as well as multiple sites of oxidation of the same peptide were identified. The high dynamic range of analysis allowed single residue resolution of the site of modification for peptides spanning an intensity range of 3 orders of magnitude and is comparable to the previous reports.[17,18] Also, the introduction of the TIMS technology allows an additional parameter, namely, CCS. Although this parameter has not been widely implemented for peptide identifications, there is an avenue for the simplification/automation of data processing. It might be useful for combining several LC−MS runs, which subsequently facilitates in an unambiguous matching of peptides bearing the same modifications. To our knowledge, the results represent a highly informative FPOP data set from the viewpoint of the number of identified modifications at single residue resolution. Identification of maximum possible sites of modification in FPOP experiments secures precise footprinting information of a protein surface. Such comprehensive analysis should result in more reliable data for structural characterization of proteins and protein−ligand complexes.

EXPERIMENTAL SECTION

Chemicals and Materials. Human haptoglobin (Hp) phenotype 1-1 was purchased from Sigma Aldrich. PNGase F was from New England BioLabs Inc. A trypsin/Lys-C mass spectrometry grade protease mixture was obtained from

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

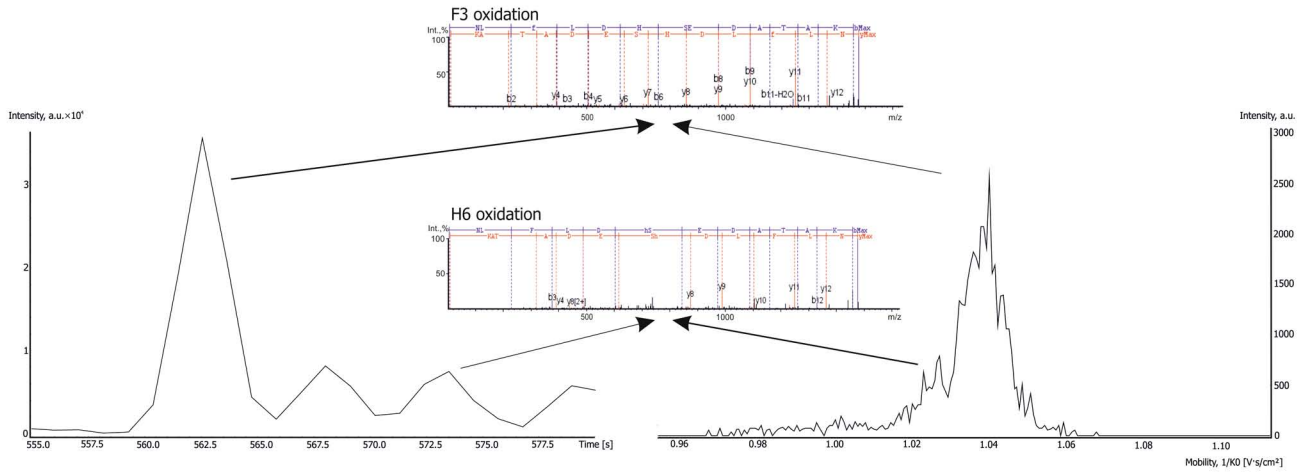


Figure 6. Separation of two forms of the peptide with sequence NLFLDHSEDATAK appearing doubly charged at m/z 738.84 using ion mobility. Left side, chromatogram; right side, mobilogram.

Table 3. List of Detected Modifications in the KQLVEIEK Peptide[a]

−10lgP	z	RT	$1/k_0$ range	intensity	oxidation site	AScore
38.14	2	2.97	0.8442−0.8628	147,992	L3	26.31
41.03	2	3.79	0.8345−0.8531	1224	V4	30.46
36.83	2	5.83	0.8286−0.8472	16,835	K8	26.31
43.59	2	6.12	0.8487−0.8673	25,697	K1	36.05
38.07	2	7.15	0.8717−0.8903	5250	I6	32.28
53.33	2	6.31	0.8583−0.8769	1,479,460	NA	NA

[a]NA, not applicable.

Promega Corporation. Additional chemicals reported in this article were purchased in the highest available purity from Merck.

SDS-PAGE Electrophoresis. The purity of purchased Hp was checked by SDS electrophoresis as follows. Hp (5 μg) was diluted in 15 μL of 150 mM ammonium acetate and mixed with 5 μL of 4× LDS (lithium dodecyl sulfate) sample buffer, including 400 mM dithiothreitol. Afterward, the sample was boiled for 5 min and loaded in a 10-well pre-casted NuPAGE Bis-Tris gradient gel of 4−12% (Thermo Fisher Scientific) placed in an electrophoretic apparatus filled with SDS-MES running buffer. After that, one well was loaded with 5 μL of pre-stained Protein Standard SeeBlue Plus2 (Thermo Fisher Scientific). Electrophoresis was run for 35 min under 200 V. The final gel was processed with a standard Coomassie Blue staining and destaining protocol.

Fast Photochemical Oxidation of Proteins (FPOP). The FPOP experiment was carried out in a continuous capillary flow system composed of two syringe pumps (New Era, models NE-1000 and NE-4000). Prior to the experiment, solutions of 0.27 mg/mL Hp, 7.5 mM H_2O_2, and 75 mM L-methionine were prepared in degassed 150 mM ammonium acetate buffer (pH 6.8). After that, three glass syringes (Hamilton) for 250, 500, and 1000 μL for the samples, H_2O_2, and L-methionine, respectively, connected with quartz capillaries (Polymicro Technologies), MicroTees (Upchurch Scientific) were filled, and pumps were started. Flow rates were set to 10, 20, and 20 μL/min for Hp, H_2O_2, and L-methionine, respectively. For generation of hydroxyl radicals, a 248 nm KrF laser (Coherent, COMPex50) was used. The laser beam was focused on a quartz capillary with an i.d. of 75 μm where a transparent window (approximately 6.5 mm) was formed by removal of the polyimide coating. Subsequently, Hp mixed with H_2O_2 was subjected to one laser shot (15 Hz, 20 ns pulse duration, 2.24 mJ/mm^2 radiant exposure, and exclusion volume was 16%) followed by mixing with 75 mM L-methionine, which served as a radical scavenger.

Protein Digestion and LC−MS/MS Analysis. Samples collected after the FPOP experiment were subjected to a reduction of cysteines with 20 mM tris(2-carboxyethyl)-phosphine, for 20 min, at 56 °C followed by alkylation with 20 mM iodoacetamide, for 20 min, at 25 °C in the dark. Subsequently, samples were twice diluted with 100 mM 4-ethylmorpholine buffer (pH 8.5):acetonitrile (ACN) (90:10 v/v) followed by overnight deglycosylation with PNGase F (protein:enzyme ratio, 1:20). Afterward, trypsin/Lys-C was added (protein:enzyme ratio, 1:20) and samples were digested for 8 h at 37 °C. Digestion was stopped by addition of TFA to a final concentration of 0.1%, and the resulting digest was subsequently dried by a SpeedVac (Eppendorf).

Samples were analyzed on an ultrahigh pressure nanoflow chromatography system (nanoElute, Bruker Daltonics) coupled to a trapped ion mobility quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonics) via a nanoelectrospray ion source (Captive Spray Source, Bruker Daltonics). Peptides (20 or 40 ng, respectively) were directly loaded and separated on an analytical column (25 cm × 75 μm, C18, 1.6 μm) (Aurora Column, Ion Opticks, Australia). Peptides were eluted using 2% ACN/0.1% formic acid as mobile phase A at a flow rate of 400 nL/min and 21 minute-long gradient with liner increase of acetonitrile to 37% (the mobile phase B was ACN/0.1% formic acid) at a 50 °C column oven temperature. The eluting peptides were interrogated by an MS acquisition method recording spectra from 100 to 1700 m/z and ion mobility scanned from 0.6 to 1.6 Vs/cm^2. The method consisted of a TIMS survey scan of

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

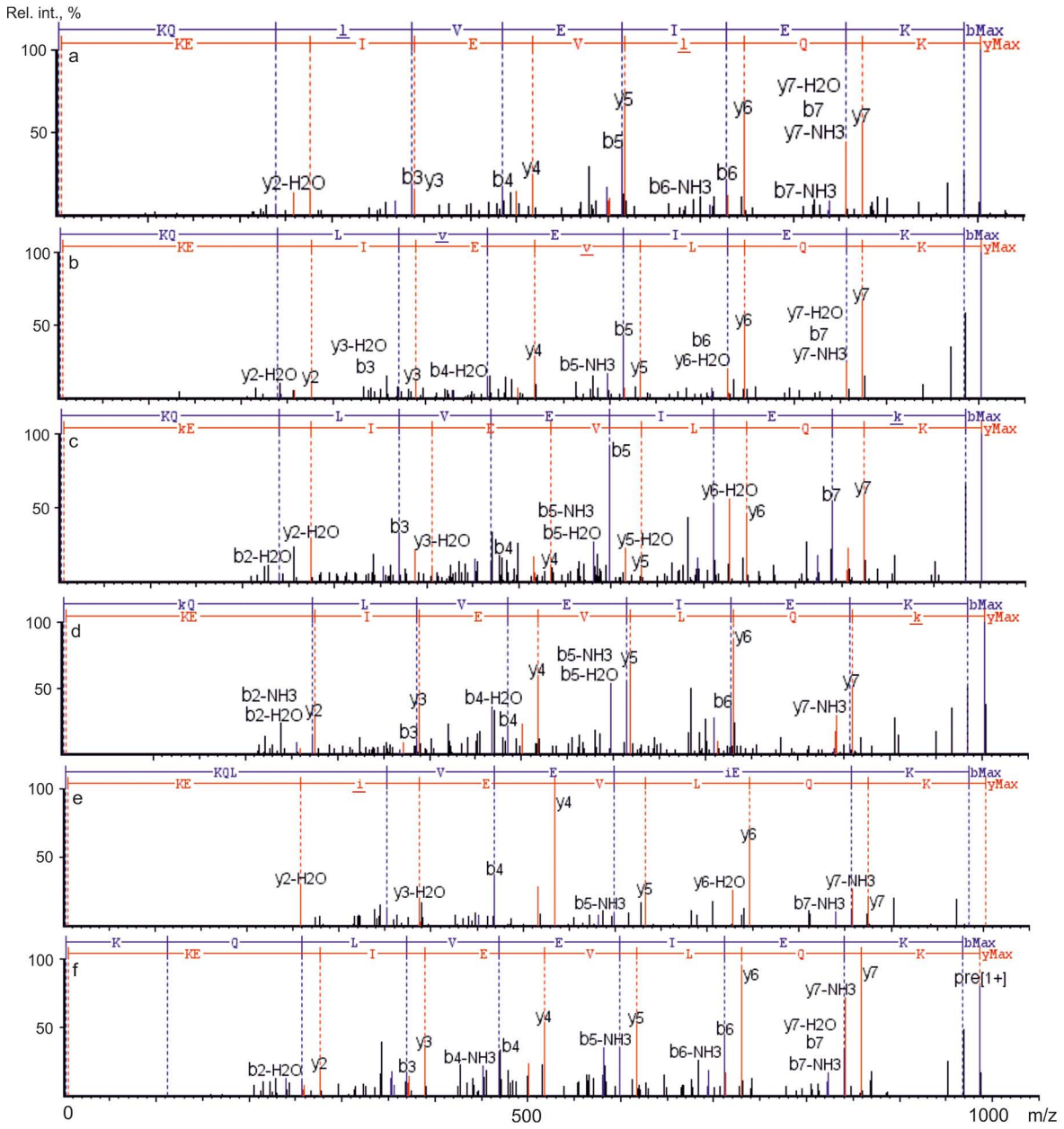


Figure 7. Fragmentation spectra of the peptide KQLVEIEK appearing doubly charged at m/z 501.78: (a) L3, (b) V4, (c) K8, (d) K1, (e) I6, and (f) without modification. Modified residues are labeled with underlined lowercase letters.

150 ms followed by six PASEF MS/MS scans, each 150 ms for ion accumulation and ramp time. The total cycle time was 1.08 s. The target intensity was 40,000, the intensity threshold was 1000, and singly charged peptides with m/z < 800 were excluded by an inclusion/exclusion polygon filter applied within the ion mobility over m/z heatmaps. Precursors for data-dependent acquisition were fragmented with an ion mobility-dependent collision energy, which was linearly increased from 20 to 59 eV.

The raw data were deposited to a ProteomeXchange Consortium via a PRIDE[19] partner repository with a data set identifier PXD020509.

Data Analysis. The acquired data were searched against a database containing the sequence of haptoglobin (deglycosylated form) supplemented with sequences of common contaminants (Max Planck Institute of Biochemistry, Martinsried, Germany) using PEAKS X+ software (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Precursor ion tolerance was set at 10 ppm, and the mass tolerance for MS/MS fragment ions was set at 0.05 Da. Carbamidomethylation of cysteine and commonly observed FPOP modifications[6] (Supporting Information, Table S1) were considered as variable modifications. Peptide-spectrum matches were filtered by peptide −10lgP scores ≥15, and the threshold to a localization score assigned to modifications was set at ≥20. FDR was set to 5%. Intensities of modified peptides were determined using Compass DataAnalysis v.5.2 (Bruker Daltonics).

To calculate the surface-accessible surface area (SASA) of Hp 1-1, respective coordinates were extracted from the model PDB 4WJG. The calculation was accomplished by NACCESS 2.1.1 software using the rolling probe algorithm.[20] A probe radius was set to 2.8 Å to mimic the composition of the H_2O_2 molecule.[21]

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

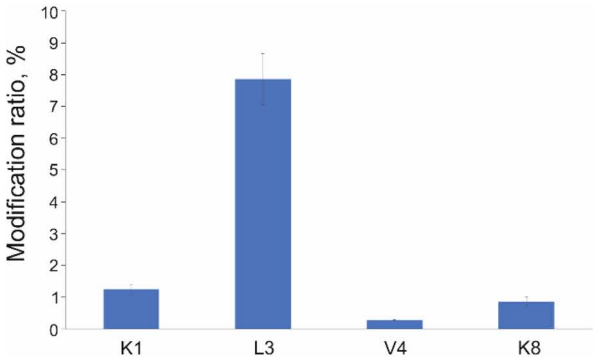


Figure 8. Reproducibility of FPOP quantitative analysis using the timsTOF Pro mass spectrometer. An oxidation of the KQLVEIEK peptide at the residue L3 represents highly abundant modification, K1 and K8 are middle abundant modifications, and V4 is a low abundant modification. Error bars represent standard deviation from three replicates.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c00732.

FPOP modifications (Table S1); list of Hpα and Hpβ peptides with a certain modification (Table S2); list of peptides separated using the ion mobility cell (Table S3); SDS-PAGE electrophoresis of haptoglobin (Figure S1); peptide map of haptoglobin (20 ng) (Figure S2); fragmentation spectra of the peptide NANFKFTDHLK (m/z 444.8847) (Figure S3); fragmentation spectra of the peptide 203−215 (m/z 738.8416) with RT 7.7 and 9.4 (Figure S4); fragmentation spectra of the peptide 271−277 (m/z 825.3568) with RT 5.6 and 6.5 (Figure S5); fragmentation spectra of the peptide 392−401 (m/z 610.3093) with RT 9.7, 10.4, and 12.4 (Figure S6); analysis reproducibility (Figure S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Dmitry S. Loginov − *Institute of Microbiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic; Orekhovich Institute of Biomedical Chemistry, 119191 Moscow, Russia;* orcid.org/0000-0002-4164-6145; Email: dmitry.loginov@biomed.cas.cz

Petr Novak − *Institute of Microbiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic;* orcid.org/0000-0001-8688-529X; Email: pnovak@biomed.cas.cz

Authors

Jan Fiala − *Faculty of Science, Charles University, 128 20 Prague, Czech Republic*

Josef Chmelik − *Institute of Microbiology, The Czech Academy of Sciences, 14220 Prague, Czech Republic*

Peter Brechlin − *Bruker Daltonik GmbH, 28359 Bremen, Germany*

Gary Kruppa − *Bruker s.r.o., 619 00 Brno, Czech Republic*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c00732

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): Two of the authors, Peter Brechlin and Gary Kruppa are employees of Bruker Daltonics. All experiments have been generated in a collaboration of Petr Novak laboratory and Bruker Daltonics, where Bruker provided the access to a timsTOF Pro mass spectrometer.

■ ACKNOWLEDGMENTS

This work was mainly supported by the Czech Science Foundation (grant number 19-16084S), the Ministry of Education of the Czech Republic (program "NPU II" project LQ1604), the European Commission H2020 (EPIC-XS - grant agreement ID: 823839), and, in part, by the Czech Academy of Sciences (RVO61388971). We thank Dr. Filip Dycka (University of South Bohemia, Czech Republic) for providing access to the PEAKS software, and we acknowledge the Centre of Molecular Structure Core Facility at BIOCEV, a facility funded by the European Regional Development Funds (CZ.1.05/1.1.00/02.0109 BIOCEV) and supported by the Czech Infrastructure for Integrative Structural Biology (structural mass spectrometry CF - LM2018127 CIISB for CMS BIOCEV funded by MEYS CR).

■ REFERENCES

(1) Aye, T. T.; Low, T. Y.; Sze, S. K. Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 5814−5822.

(2) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057−2063.

(3) Wang, L.; Chance, M. R. Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem.* **2011**, *83*, 7234−7241.

(4) Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, 4355.

(5) Johnson, D. T.; Di Stefano, L. H.; Jones, L. M. Fast Photochemical Oxidation of Proteins (FPOP): A Powerful Mass Spectrometry−Based Structural Proteomics Tool. *J. Biol. Chem.* **2019**, 11969−11979.

(6) Xu, G.; Chance, M. R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007**, *107*, 3514−3543.

(7) Liu, X. R.; Zhang, M. M.; Zhang, B.; Rempel, D. L.; Gross, M. L. Hydroxyl-Radical Reaction Pathways for the Fast Photochemical Oxidation of Proteins Platform As Revealed by 18O Isotopic Labeling. *Anal. Chem.* **2019**, *91*, 9238−9245.

(8) Cornwell, O.; Bond, N. J.; Radford, S. E.; Ashcroft, A. E. Long-Range Conformational Changes in Monoclonal Antibodies Revealed Using FPOP-LC-MS/MS. *Anal. Chem.* **2019**, *91*, 15163−15170.

(9) Liu, X. R.; Rempel, D. L.; Gross, M. L. Protein Higher-Order-Structure Determination by Fast Photochemical Oxidation of Proteins and Mass Spectrometry Analysis. *Nat. Protoc.* **2020**, *15*, 3942−3970.

(10) Silveira, J. A.; Ridgeway, M. E.; Park, M. A. High Resolution Trapped Ion Mobility Spectrometery of Peptides. *Anal. Chem.* **2014**, *86*, 5624−5627.

(11) Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M. Parallel Accumulation-Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by

Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14*, 5378−5387.
(12) Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Räther, O.; Mann, M. Online Parallel Accumulation−Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **2018**, *17*, 2534−2545.
(13) Sandow, J. J.; Infusini, G.; Dagley, L. F.; Larsen, R.; Webb, A. I. Simplified High-Throughput Methods for Deep Proteome Analysis on the Tims TOF Pro. *bioRxiv* **2019**, 657908.
(14) Stødkilde, K.; Torvund-Jensen, M.; Moestrup, S. K.; Andersen, C. B. F. Structural Basis for Trypanosomal Haem Acquisition and Susceptibility to the Host Innate Immune System. *Nat. Commun.* **2014**, *5*, 1−8.
(15) Xu, C. F.; Chen, Y.; Yi, L.; Brantley, T.; Stanley, B.; Sosic, Z.; Zang, L. Discovery and Characterization of Histidine Oxidation Initiated Cross-Links in an IgG1 Monoclonal Antibody. *Anal. Chem.* **2017**, *89*, 7915−7923.
(16) Cornwell, O.; Radford, S. E.; Ashcroft, A. E.; Ault, J. R. Comparing Hydrogen Deuterium Exchange and Fast Photochemical Oxidation of Proteins: A Structural Characterisation of Wild-Type and ΔN6 B2-Microglobulin. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2413−2426.
(17) Chea, E. E.; Jones, L. M. Modifications Generated by Fast Photochemical Oxidation of Proteins Reflect the Native Conformations of Proteins. *Protein Sci.* **2018**, *27*, 1047−1056.
(18) Zhang, M. M.; Rempel, D. L.; Gross, M. L. A Fast Photochemical Oxidation of Proteins (FPOP) Platform for Free-Radical Reactions: The Carbonate Radical Anion with Peptides and Proteins. *Free Radical Biol. Med.* **2019**, *131*, 126−132.
(19) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442−D450.
(20) Hubbard, S. J.; Thornton, J. *NACCESS*. Department of Biochemistry and Molecular Biology, University College: London 1993.
(21) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379.

NOTE ADDED AFTER ISSUE PUBLICATION

The affiliations of the first two authors were incorrectly swapped in the original version of this article. The revised version was published to the Web on May 28, 2021. An Addition & Correction was published in volume 6, issue 22.

https://doi.org/10.1021/acsomega.1c00732
ACS Omega 2021, 6, 10352−10361

ARTICLE III

Loginov D., Fiala J., Brechlin P., Kruppa G., Novák P.

Hydroxyl radical footprinting analysis of a human haptoglobin-hemoglobin complex.

Biochim. Biophys. Acta., **accepted**

My contribution: Conducting experiments, LC-MS/MS analysis, writing methodological part in manuscript and creating figures

Hydroxyl radical footprinting analysis of a human haptoglobin-hemoglobin complex.

Dmitry S. Loginov[a,b,*], Jan Fiala[a,c], Peter Brechlin[d], Gary Kruppa[e], Petr Novak[a,*]

a – BioCeV - Institute of Microbiology of the CAS, Prumyslova 595, CZ-252 50 Vestec, Czech Republic

b – Orekhovich Institute of Biomedical Chemistry, Pogodinskaja str. 10, 119191, Moscow, Russia

c – Faculty of Science, Charles University, Hlavova 8, CZ-128 20 Prague, Czech Republic

d – Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany

e – Bruker s.r.o., Prazakova 60, 619 00 Brno, Czech Republic

* – To whom correspondence should be addressed: dmitry.loginov@biomed.cas.cz (D. Loginov); pnovak@biomed.cas.cz (P. Novak)

Abstract

Methods of structural mass spectrometry have become more popular to study protein structure and dynamics. Among them, fast photochemical oxidation of proteins (FPOP) has several advantages such as irreversibility of modifications and more facile determination of the site of modification with single residue resolution. In the present study, FPOP analysis was applied to study the hemoglobin (Hb) – haptoglobin (Hp) complex allowing identification of respective regions altered upon the complex formation. Oxidative modifications were precisely localized on specific residues using a timsTOF Pro mass spectrometer. The data allowed determination of amino acids directly involved in Hb – Hp interactions and those located outside of the interaction interface yet affected by the complex formation. Data are available via ProteomeXchange with identifier PXD021621.

Key words: FPOP, footprinting, haptoglobin, hemoglobin, timsTOF Pro

1. Introduction

Currently, methods of structural mass spectrometry have become a powerful tool in the field of structural biology facilitating studies of protein folding and interactions. In this field, fast photochemical oxidation of proteins (FPOP) introduced in 2005 [1] provides valuable data related to a structural dynamics of proteins. This method is based on an irreversible covalent modification of amino acids side chains by hydroxyl radicals, which can be generated by a variety of methods including synchrotron radiolysis of water, electrochemistry, or laser photolysis of H_2O_2 [2]. The broad reactivity of hydroxyl radicals results in possible modifications of 19 amino acids, and 14 of these modifications are readily detected by mass spectrometry so that FPOP-MS analysis provides a rich and informative dataset related to chemical footprinting of the surfaces of proteins [3].

Reactivity of amino acids with $^{\bullet}OH$ varies by up to two orders of magnitude [3] meaning some difficulties may be encountered in detection of modified peptides with low relative abundance. Thus, developments in the instrumentation, e.g. increasing sensitivity of mass spectrometers, and improvements in peptide separation methods, etc., will allow more informative datasets to be obtained. From this point of view, a new mass spectrometer, the timsTOF Pro from Bruker Daltonics has features that should be very helpful. An integrated trapped ion mobility spectrometry (TIMS) adds a third dimension increasing the capacity for peptide separation [4], and the "Parallel Accumulation – Serial Fragmentation" (PASEF) method secures fast, reliable, and very sensitive fragmentation data from multiple precursors [5,6].

Hemoglobin (Hb) is one of the key proteins of the respiratory system, which transports oxygen and participates in the detoxification of reactive oxygen and nitrogen species [7]. Hb forms a tetramer by symmetric pairing of dimers consisting of the α- and β-globins, where each subunit contains a prosthetic heme molecule [8]. Erythrocyte-encapsulated Hb serves as a protector, but being released from them, it dissociates into αβ dimers and transforms into a very toxic protein due to the reactive properties of the heme [7]. To prevent possible toxic events an organism synthesizes an acute phase protein haptoglobin (Hp), which has the highest binding affinity for Hb [9]. In its turn, the Hb – Hp complex is recognized by the macrophage receptor CD163, which scavenges Hb [10]. The simplest Hp form is Hp 1-1 composed from two α- and two β-chains linked by disulphide bonds [11]. Currently, two X-ray structures of Hb – Hp complex are accessible in the PDB database. However, one structure represents complex of porcine proteins with a resolution 2.9 Å [12], and the second is a multimeric complex of human Hb – Hp with a receptor from *Trypanosoma brucei brucei* [13]. So far, there is not any high-resolution structural model of human Hb – Hp complex alone. Based on the structural data several amino acids participating in the Hb – Hp interactions have been indicated in porcine complex [12,14], yet the finest molecular details remains unclear for the human protein complex [15]. Thus, we applied FPOP analysis to study human Hb – Hp interactions in more details.

2. Materials and methods

Lyophilized human haptoglobin phenotype 1-1 (Hp) and human α,β haemoglobin (Hb) were purchased from Sigma Aldrich. Deglycosylation enzyme PNgase F was from New England BioLabs Inc. For digestion Typsin/Lys-C mass spectrometry grade proteases

mixture from Promega Corporation was used. Additional chemicals reported in this article were purchased in the highest available purity from Merck.

2.1 Fast photochemical oxidation of proteins (FPOP)

Before the FPOP analysis the purity of the proteins was confirmed by SDS-PAGE electrophoresis (Supplementary fig. 1). Solutions of 0.27 mg/ml Hp, 0.27 mg/ml Hb, 7.5 mM H_2O_2 and 75 mM L-methionine were prepared in degassed 150 mM ammonium acetate (pH 6.8). The protein complex was formed by mixing Hp and Hb stock solutions in a volumetric ratio 1:1. The FPOP experiment was carried out in a continuous capillary flow system described in [16]. The system is composed of two syringe pumps (New Era, models NE-1000 and NE-4000) and three glass syringes (Hamilton) connected with quartz capillaries (Polymicro Technologies) and two MicroTees (Upchurch Scientific). Before the experiment, 250, 500 and 1000 µl syringes were filled with the protein sample, 7.5 mM H_2O_2 and 75 mM L-methionine, respectively. Flow rates were set to 10 ul/min, 20 ul/min and 20 ul/min, respectively. For the generation of hydroxyl radicals, a 248 nm KrF laser (Coherent, COMPex50) was used. The laser beam was focused on a quartz capillary with i.d. of 75 µm where a transparent window (approx. 6.5 mm) was formed by removal of the polyimide coating. Subsequently, the protein sample mixed with H_2O_2 was irradiated with one laser shot (15 Hz, 20 ns pulse duration, 2.24 mJ/mm^2 radiant exposure) followed by quenching of hydroxy radicals with 75 mM L-methionine.

2.2 Protein digestion and LC-MS/MS analysis

Samples collected after the FPOP experiment were subjected to a reduction of cysteines with 20 mM Tris(2-carboxyethyl)phosphine, for 20 min, at 56 °C, followed by alkylation with 20 mM iodoacetamide for 20 min at 25 °C in the dark. Samples were then two times diluted with 100 mM 4-ethylmorpholine buffer (pH 8.5): acetonitrile (ACN) (90:10 v/v), followed by overnight deglycosylation with PNGase F (protein:enzyme ratio 1:20). Afterwards, trypsin/Lys-C was added (protein:enzyme ratio 1:20) and samples were digested for 8 hours at 37 °C. Digestion was stopped by addition of TFA to a final concentration of 0.1% and tryptic peptides were subsequently dried by SpeedVac (Eppendorf).

Samples were analyzed on an ultra-high pressure nanoflow chromatography system (nanoElute, Bruker Daltonics) coupled to a quadrupole time-of-flight mass spectrometer with a trapped ion mobility (TIMS) front end (timsTOF Pro, Bruker Daltonics) via a nano-electrospray ion source (Captive Spray Source, Bruker Daltonics). By microliter pickup, peptides (in amount corresponding to the 40 ng of digested proteins) were directly loaded and separated on an analytical column 25 cm x 75 µm, C18, 1.6 µm (Aurora Column, Ion Opticks, Australia). Peptides were eluted using 2 % ACN, 0.1% formic acid as mobile phase A at a flow rate of 400 nl/min, applying a 21 min gradient of 2-37% Acetonitrile by mobile phase B: ACN, 0.1% formic acid at 50 °C column oven temperature. The eluting peptides were interrogated by an MS acquisition method recording spectra from 100-1,700 m/z and ion mobility scanned from 0.6 to 1.6 Vs/cm^2. The method consisted of a TIMS survey scan of 150 ms followed by 6 PASEF MS/MS scans, each 150 ms for ion accumulation and ramp time (ion accumulation is carried out in parallel, ions are accumulated for the next scan in the accumulation region of the TIMS device, while ions are being scanned out of the analyzer region). Total cycle time was 1.08 seconds. Target intensity was 40,000, intensity threshold 1000, singly charged peptides with $m/z < 800$ were excluded by an inclusion/exclusion

polygon filter applied within the ion mobility – m/z heatmap. Precursors for data-dependent acquisition were fragmented with an ion mobility-dependent collision energy, which was linearly increased from 20 to 59 eV.

The raw data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD021621.

2.3 Data analysis

Acquired data were searched against a database containing the sequences of Hb and Hp (including deglycosylated form of β subunit) supplemented with sequences of common contaminants using PEAKS X+ software (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Precursor ion tolerance was set at 10 ppm, and the mass tolerance for MS/MS fragment ions was set at 0.05 Da. Carbamidomethylation of cysteine and commonly observed FPOP modifications [3] were considered as variable modifications (Table 1).

Table 1. List of FPOP modifications.

Modification	Targets
FPOP oxidation / + 15.994915	M, W, Y, F, H, L, I, V, P, R, K
FPOP dioxydation / + 31.989828	C, M, W, Y, F
FPOP trioxydation / + 47.984745	C, W
FPOP carbonyl / + 13.979265	L, I, V, P, R, E, Q, H
FPOP Met-32Da / - 32.008457	M
FPOP His+5Da / 4.9789	H
FPOP His-10Da / -10.031968	H
FPOP His->Asp / -22.03197	H
FPOP His->Asn / -23.01598	H
FPOP Arg_deguanidination / - 43.053432	R
FPOP decarboxylation / - 30.010565	Q, N
FPOP ST_carbonyl / - 2.01565	S, T

Peptide-spectrum matches were filtered by peptide -10lgP scores ≥ 15, and threshold to a localization score assigned to modifications was set ≥ 20. FDR was set to 5 %. Intensities of peptides were determined using Compass DataAnalysis v.5.2 (Bruker Daltonics).

3. Results and discussion

Studying the molecular details of the human Hb – Hp interaction is a challenging task. Although both proteins have been subjected to various investigations and two 3D structures with resolution around 3 Å of their complex have been published [12,13], none of them represent a high resolution structural model of human Hb – Hp complex. Thus, methods of structural mass spectrometry might be useful in refining the proposed structural models. In the present study, FPOP was the method of choice, because of the deglycosylation step in the sample preparation, which is hardly compatible with HDX MS, an another common method used for protein structural studies [2]. Data from single proteins and their complexes were compared to determine regions and residues involved in the Hb – Hp interaction or altered by it.

Using a bottom-up approach a sequence coverage of Hpα and Hpβ was 100 % for the "free" protein and in the complex with Hb. Also, full sequence coverage was achieved for

Hbα and Hbβ in the complex. For "free" Hb coverage was 59 % and 97 % for α and β subunits, respectively (Supplementary fig. 2-5).

Protein digests were measured using a recently introduced mass spectrometer, the timsTOF Pro with TIMS and PASEF technologies allowing an additional dimension of peptide separation (ion mobility) and fragmentation of a high number of precursors without loss of sensitivity [6]. Data were processed using PEAKS X+ software. A benefit of using PEAKS X+ is the possibility to introduce a high number of peptide modifications with reasonable computational time. Overall, 13 different modifications were included in the data search, including those occurring at multiple residues, like oxidation (+16 Da) of Phe, His, Ile, Lys, Leu, Met, Pro, Arg, Val, Trp, Tyr and +14 Da mass shift at Glu, His, Ile, Leu, Pro, Gln, Arg, and Val. Thus, we were able to get a very informative and rich dataset of identified peptides bearing up to three modifications (Table 2).

Table 2. Number of identified peptides by PEAKS X+ software.

	Hbα			Hbβ			Hpα			Hpβ		
	All	Mod.	Cert.	All	Mod.	Cert.	All	Mod.	Cert.	All	Mod.	Cert.
Hb	149	78	34	218	142	50	-	-	-	-	-	-
Hp	-	-	-	-	-	-	89	29	2	282	116	58
Hb-Hp	294	144	49	335	184	52	169	59	2	429	184	61

All – all identified peptides; Mod. – number of modified peptides with a localization score ≥ 20; Cert. – number of modified peptides with at least one definite modified residue.

The high number of modified peptides arises from multiple possible modifications of the same residues. For example, six different modifications of His and Trp in mono-, di- and trioxidized states were found (Supplementary fig. 2-5). Taken together, these data allow precise footprinting of protein surfaces upon their interaction.

Quantitative analysis of modified peptides allowed the discovery of protein regions affected upon Hb – Hp complex formation (Fig. 1). Hpα was more oxidized (less protected) in the complex, including the C and N termini, while no significantly protected regions were observed. Rearrangements in the Hpβ structure upon the interaction with Hb resulted in protection of the N terminus and the region 271-311, and more exposure to oxidation of the C terminus and the region 203-270. On the contrary, protection of the C terminus and oxidation of the N terminus were observed in the case of Hbα. For Hbβ a more oxidized protein region upon complex formation was residues 105-146, including the N terminus. It should be mentioned that oxidation of the middle parts of Hbβ was detected only upon the complex formation (Fig 1c and 1d).

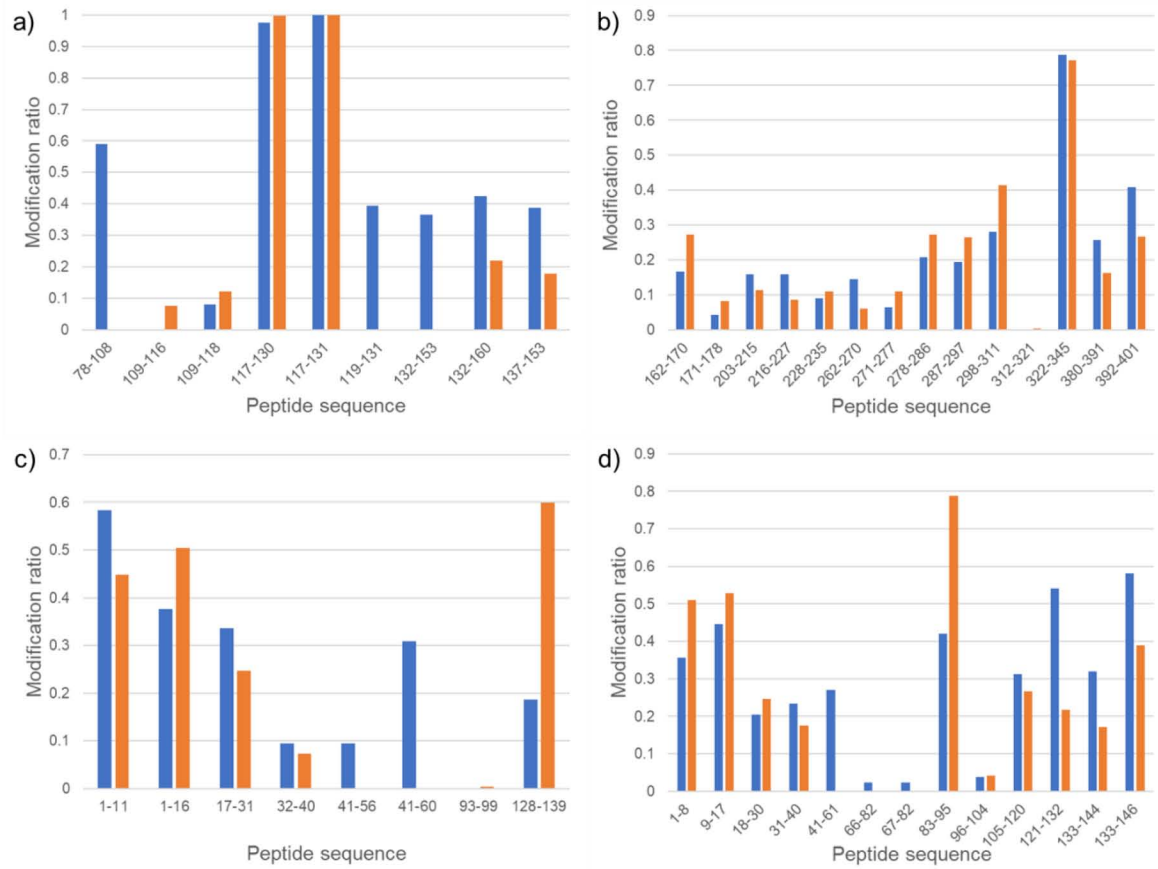


Figure 1. Extent of modification of peptides from Hb and Hp (Blue bars – Hb – Hp complex, orange bars – "free" protein). a – Hpα, b – Hpβ, c – Hbα, d – Hbβ.

One of the advantages of FPOP analysis over HDX is that single residue resolution is easier to achieve, as it is not necessary to resort to ETD or other non-ergodic fragmentation schemes to avoid scrambling of the labels. For this analysis we selected peptides with certainly determined modified residues and S/N ratio > 900 to obtain unambiguous data. Overall, an oxidation rate was determined for 2, 37, 27 and 32 residues in Hpα, Hpβ, Hbα and Hbβ, respectively (Fig. 2). Both certain modifications identified for Hpα corresponded to the trioxidation of cysteines. Among them, Cys111, the residue engaged in a disulfide bridge formation, that connects two αβHp dimers [12], was found to be susceptible to oxidation in the complex at a higher rate.

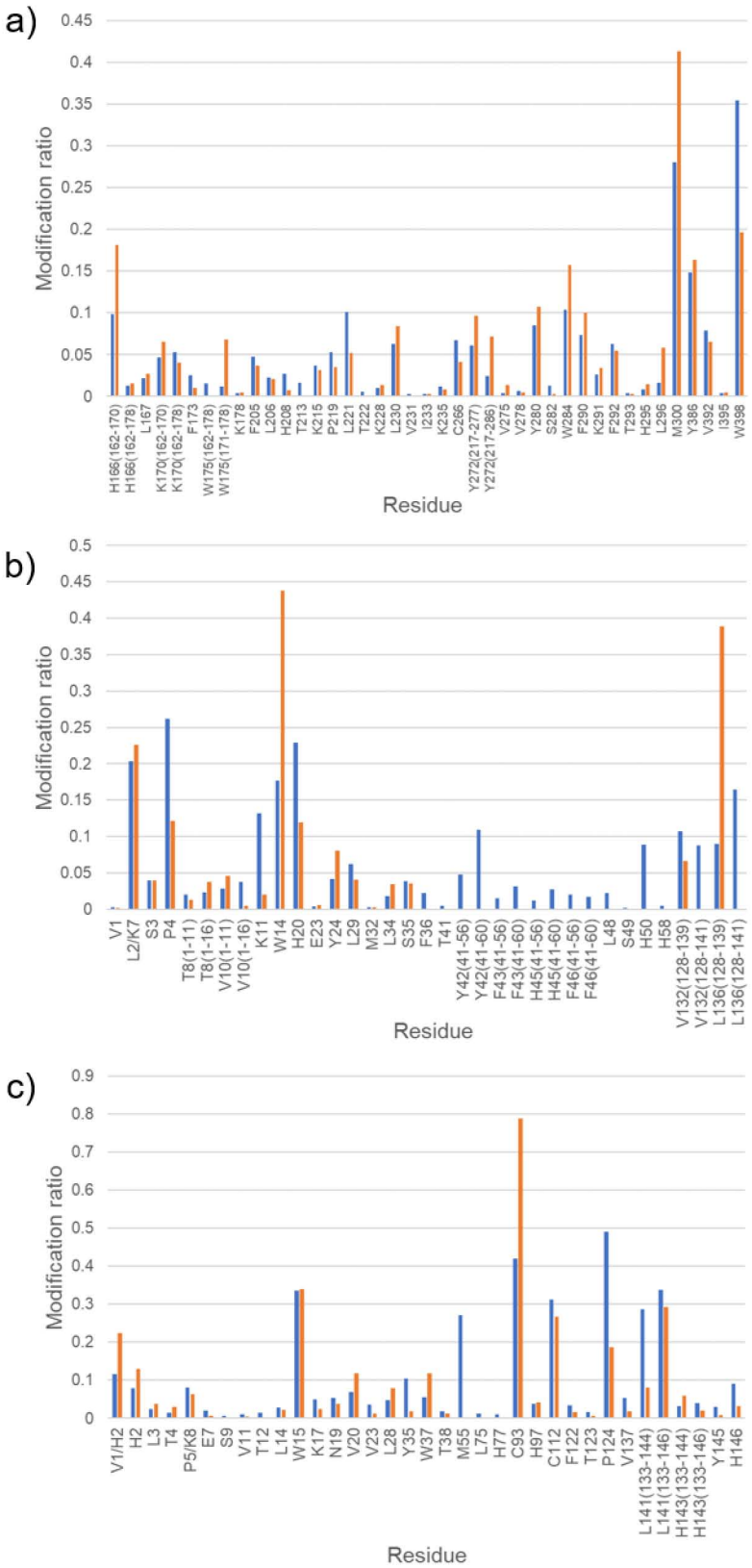


Figure 2. Modification ratios at the residue level (Blue bars – Hb – Hp complex, orange bars – "free" protein). a – Hpβ, b – Hbα, c – Hbβ. For the residues with multiple modifications, like His, the sum of all detected modifications contributing to the overall

oxidation extent is presented (see Supplementary Table 1 for the respective values for the each modification).

Previously, several sites of Hb oxidation by H_2O_2 have been described [17,18]. The proposed hot spots included Tyr42 in the α subunit, Cys93 and Tyr145 in the β subunit. Also, His2, Trp15, Pro36, Trp37, Met55 and Cys112 in the Hbβ were found to be minor oxidative sites. Protection (decreased rate or absence of an oxidation) of all these sites upon Hb – Hp complex formation has also been shown [18]. Evidence for modification of all these residues but βPro36 were found during our FPOP analysis. Because of differences in the reaction conditions we observed modification of the neighboring residue βTyr35 which is a more favorable amino acid for the oxidation by $^{\bullet}OH$ [3] , instead βPro36.

Evidence for strong protection of these Hb residues against oxidation upon interaction with Hp found in the previous study was not confirmed by our data. αTyr42 and βMet55 were found to be readily oxidized in the complex. None of the residues previously reported to be susceptible to oxidation by H_2O_2 were found to be completely protected upon Hb – Hp complex formation (Fig. 2). This discrepancy could be explained by the conditions of the radical generation. In Pimenova et al radicals were generated by Hb itself, thus no oxidation was found in the case of Hb – Hp complex [18]. In the present study radicals were generated by laser induced homolytic cleavage of H_2O_2.

Nine residues found to be modified during our FPOP analysis have been shown to be potentially involved in Hb – Hp interactions (Hb: αVal1, βTrp37, βHis97 and βTyr145 ; Hp: βLeu167, βLeu206, βPhe290, βLys291 and βHis295) [12]. No significant changes in the modification ratio of these residues in Hpβ were found (Fig. 2a). However, the overall oxidation level of the respective peptides, except 203-215, was lower in the complex, meaning the respective protein parts are protected due to Hb – Hp interactions (Fig. 1b). Residues 203-215 represents an unstructured region located outside of the interaction interface. Thus, its higher oxidation rate might be explained by some conformational changes in the complex favorable to the subsequent oxidation of amino acids side chains by hydroxyl radicals. Two Hb residues, namely αVal1 and βHis97, were oxidized at the same level as in the "free" protein (Fig. 2).

Previously βTrp37 and βPhe292 in Hb and Hp, respectively, have been proposed as contributing the most to the binding energy of Hb – Hp complex, so called hot spots, and interacting with each other based on the protein-protein docking calculation using structure of porcine Hb – Hp complex as the template [14]. The last was not confirmed by our data. The Hb βTrp37 residue was significantly protected upon the complex formation, whereas Hp βPhe292 was oxidized at the same level (Fig. 2). Thus, Hb βTrp37 is indeed involved in intermolecular interactions, but not with Hp βPhe292. This finding is in agreement with the X-ray structural model of human complex, where the Cα-Cα distance between these residues is 18.5Å making the interaction unlikely [13]. Further Phe292 is located on the surface of Hpβ in a near proximity of two complex glycan moieties (Asn184 and Asn207) protecting it from the possible complex formation [19].

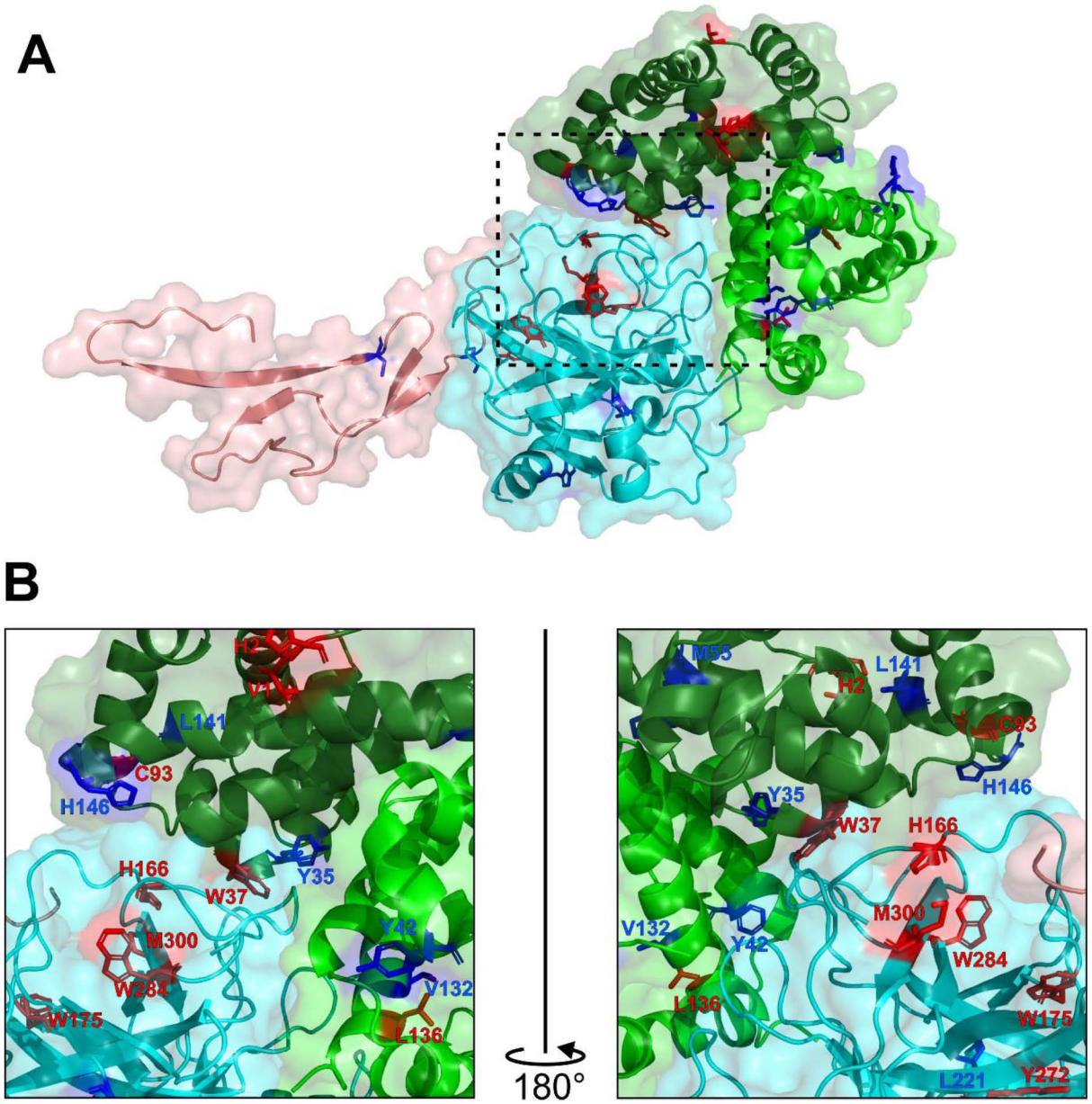


Figure 3. Structural model of Hb – Hp complex with highlighted residues oxidized by the FPOP experiment. a – Overview of a tetrameric complex composed of α,βHp and α,βHb subunits. b – Zoom of interaction interface. Red residues were found to be less oxidized (protected) upon the formation of complex, showing evidence of the interaction interface. Blue residues represent amino acids that were more oxidized after the complex's formation. Hpα is in pink, Hpβ is in cyan, Hbα is in green and Hbβ is in dark green. Respective light colors represent solvent accessible area of the protein surface.

Analysis of oxidation extents of different residues allowed the determination of the most affected residues upon the complex formation (Fig. 3). As expected, a majority of the protected residues were located in the Hb – Hp interaction interfaces, whereas the more oxidized ones were on the protein surfaces. Differences in the oxidation rates of Hp residues were less profound than in Hb. That is not surprising because Hb is subjected to more structural changes upon Hb – Hp complex formation. The Hb residues with increased oxidation rate located on the surface of α (Pro4, Lys11, Tyr42 and His50) and β (Met55 and

Pro124) subunits might be considered as participating in the formation of Hb quaternary structure.

4. Conclusions

FPOP analysis enables an efficient structural analysis of Hb – Hp interactions. The advantage of using FPOP for studying of protein-protein interactions is the possibility to follow even minor conformational changes providing a complete dataset related to a possible interaction mechanism. The FPOP data allowed determination of regions in the protein structures affected upon the complex formation, and single residues involved in the interactions were identified. Three Hpβ residues located in the interaction interface, namely His166, Trp284 and Met300, were shown to be protected upon the complex formation. Also, lower oxidation rates were determined for αLeu137, βTrp37 and βCys94 in Hb. These data indicate involvement of the residues in the Hb – Hp interaction. The FPOP dataset also identifies Hb residues participating in the formation of its quaternary structure.

Declaration of Competing Interest

The authors declare no competing financial interest.

Supporting Information

Supplementary figure 1. SDS-PAGE electrophoresis of human α,β hemoglobin and human haptoglobin phenotype 1-1. Lane 1 – hemoglobin, lane 2 – haptoglobin, Std – protein ladders.

Supplementary figure 2. Peptide map of haptoglobin α. A – HpHb complex, B – free protein. Modifications localized on the particular residue from fragmentation spectra are indicated above the sequence line.

Supplementary figure 3. Peptide map of haptoglobin β. A – HpHb complex, B – free protein. Modifications localized on the particular residue from fragmentation spectra are indicated above the sequence line.

Supplementary figure 4. Peptide map of hemoglobin α. A – HpHb complex, B – free protein. Modifications localized on the particular residue from fragmentation spectra are indicated above the sequence line.

Supplementary figure 5. Peptide map of hemoglobin β. A – HpHb complex, B – free protein. Modifications localized on the particular residue from fragmentation spectra are indicated above the sequence line.

Supplementary table 1. List of multiple modifications detected on the same residues.

Acknowledgements

We thank Dr. Filip Dycka (University of South Bohemia, Czech Republic) for providing access to the PEAKS software and we acknowledge the Centre of molecular structure Core Facility at BIOCEV, a facility funded by European Regional Development Funds (CZ.1.05/1.1.00/02.0109 BIOCEV) and sup-ported by the Czech Infrastructure for Integrative Structural Biology (Structural mass spectrometry CF - LM2018127 CIISB for CMS BIOCEV funded by MEYS CR).

Funding sources

This work was mainly supported by the Czech Science Foundation (grant number 19-16084S), the Ministry of Education of the Czech Republic (program "NPU II" project LQ1604), European Commission H2020 (EPIC-XS - grant agreement ID: 823839) and, in part, by the Czech Academy of Sciences (RVO61388971).

References

[1] T.T. Aye, T.Y. Low, S.K. Sze, Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry, Anal. Chem. 77 (2005) 5814–5822. https://doi.org/10.1021/ac050353m.

[2] X.R. Liu, M.M. Zhang, M.L. Gross, Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications, Chem. Rev. (2020). https://doi.org/10.1021/acs.chemrev.9b00815.

[3] G. Xu, M.R. Chance, Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, Chem. Rev. 107 (2007) 3514–3543. https://doi.org/10.1021/cr0682047.

[4] J.A. Silveira, M.E. Ridgeway, M.A. Park, High resolution trapped ion mobility spectrometery of peptides, Anal. Chem. 86 (2014) 5624–5627. https://doi.org/10.1021/ac501261h.

[5] F. Meier, S. Beck, N. Grassl, M. Lubeck, M.A. Park, O. Raether, M. Mann, Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device, J. Proteome Res. 14 (2015) 5378–5387. https://doi.org/10.1021/acs.jproteome.5b00932.

[6] F. Meier, A.D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer, Mol. Cell. Proteomics. 17 (2018) 2534–2545. https://doi.org/10.1074/mcp.TIR118.000900.

[7] I.K. Quaye, Extracellular hemoglobin: The case of a friend turned foe, Front. Physiol. 6 (2015) 1–13. https://doi.org/10.3389/fphys.2015.00096.

[8] A.N. Schechter, K. Singer, H. Lehmann, W. Castle, T. Huisman, E. Jaffe, ASH 50th anniversary review Hemoglobin research and the origins of molecular medicine, Blood. 112 (2008) 3927–3938. https://doi.org/10.1182/blood-BLOOD.

[9] P. Ascenzi, A. Bocedi, P. Visca, F. Altruda, E. Tolosano, T. Beringhelli, M. Fasano, Hemoglobin and heme scavenging, IUBMB Life. 57 (2005) 749–759. https://doi.org/10.1080/15216540500380871.

[10] M. Kristiansen, J.H. Graversen, C. Jacobsen, O. Sonne, H.J. Hoffman, S.K.A. Law, S.K. Moestrup, Identification of the haemoglobin scavenger receptor, Nature. 409 (2001) 198–201. https://doi.org/10.1038/35051594.

[11] F. Polticelli, A. Bocedi, G. Minervini, P. Ascenzi, Human haptoglobin structure and function - A molecular modelling study, FEBS J. 275 (2008) 5648–5656. https://doi.org/10.1111/j.1742-4658.2008.06690.x.

[12] C.B.F. Andersen, M. Torvund-Jensen, M.J. Nielsen, C.L.P. De Oliveira, H.P. Hersleth, N.H. Andersen, J.S. Pedersen, G.R. Andersen, S.K. Moestrup, Structure of

the haptoglobinhaemoglobin complex, Nature. 489 (2012) 456–459. https://doi.org/10.1038/nature11369.

[13] K. Stødkilde, M. Torvund-Jensen, S.K. Moestrup, C.B.F. Andersen, Structural basis for trypanosomal haem acquisition and susceptibility to the host innate immune system, Nat. Commun. 5 (2014) 1–8. https://doi.org/10.1038/ncomms6487.

[14] C. Nantasenamat, V. Prachayasittikul, L. Bulow, Molecular Modeling of the Human Hemoglobin-Haptoglobin Complex Sheds Light on the Protective Mechanisms of Haptoglobin, PLoS One. 8 (2013). https://doi.org/10.1371/journal.pone.0062996.

[15] S. Tamara, V. Franc, A.J.R. Heck, A wealth of genotype-specific proteoforms fine-tunes hemoglobin scavenging by haptoglobin, Proc. Natl. Acad. Sci. U. S. A. (2020). https://doi.org/10.1073/pnas.2002483117.

[16] X.R. Liu, D.L. Rempel, M.L. Gross, Protein higher-order-structure determination by fast photochemical oxidation of proteins and mass spectrometry analysis, Nat. Protoc. 15 (2020) 3942–3970. https://doi.org/10.1038/s41596-020-0396-3.

[17] Y. Jia, P.W. Buehler, R.A. Boykins, R.M. Venable, A.I. Alayash, Structural basis of peroxide-mediated changes in human hemoglobin: A novel oxidative pathway, J. Biol. Chem. 282 (2007) 4894–4907. https://doi.org/10.1074/jbc.M609955200.

[18] T. Pimenova, C.P. Pereira, P. Gehrig, P.W. Buehler, D.J. Schaer, R. Zenobi, Quantitative mass spectrometry defines an oxidative hotspot in hemoglobin that is specifically protected by haptoglobin, J. Proteome Res. 9 (2010) 4061–4070. https://doi.org/10.1021/pr100252e.

[19] P. Pompach, K.B. Chandler, R. Lan, N. Edwards, R. Goldman, Semi-automated identification of N-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search, J. Proteome Res. 11 (2012) 1728–1740. https://doi.org/10.1021/pr201183w.

ARTICLE IV

Fojtík L., Fiala J., Pompach P., Matoušek V., Beier P., Kukačka Z., Novák P.

Fast Fluoroalkylation of Proteins Uncovers Structure and Dynamics of Biological Macromolecules.

JACS, **under review**

My contribution: Development of apparatus, conducting experiments, creating structure figures

Fast Fluoroalkylation of Proteins Uncovers Structure and Dynamics of Biological Macromolecules.

Lukáš Fojtík[†,‡], Jan Fiala[†,‡], Petr Pompach[†,‡], Václav Matoušek[¶], Petr Beier[§], Zdeněk Kukačka[*,†] and Petr Novák[*,†,‡].

[†]Institute of Microbiology of the Czech Academy of Sciences, 14220 Prague, Czech Republic

[‡]Department of Biochemistry, Faculty of Science, Charles University, 12843 Prague, Czech Republic

[¶]CF Plus Chemicals, 62100 Brno, Czech Republic

[§]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, 16610 Prague, Czech Republic

KEYWORDS: protein structure, conformational changes, radicals, trifluoromethylation, radical labeling, Togni reagents, fluoroalkylation, mass spectrometry, hypervalent iodine reagents, protein fingerprinting, epitope mapping, protein surface mapping, proteomics, interactomics, protein footprinting

ABSTRACT: Stable covalent labeling of proteins in combination with mass spectrometry has been established as a complementary technique to classical structural methods, such as X-Ray, NMR, or Cryo-EM used for protein structure determination. Although the current stable covalent labeling techniques enable to monitor protein solvent accessible areas with sufficient spatial resolution, there is still high demand for alternative, less complicated and inexpensive approaches. Here, we introduce a new covalent labeling method based on Fast Fluoroalkylation of Proteins (FFAP). The FFAP uses fluoroalkyl radicals formed by reductive decomposition of Togni reagents with ascorbic acid for labeling of proteins on a time scale of seconds. The feasibility of FFAP for effective labeling of proteins was demonstrated by monitoring differential amino acids modification of native horse heart apomyoglobin/holomyoglobin and human haptoglobin-hemoglobin complex.

A common feature of protein labeling methods in conjunction with mass spectrometry (MS) is the identification of a novel covalent bond which is formed upon treatment of the protein by a modifying agent. The underlying principle behind these methods assumes that the degree of exposure to the solvent correlates with the accessibility and reactivity of the target sites on the protein surface. Thus, the output of the MS analysis of modified proteins is the knowledge of the surface accessible and/or reactive area of the protein molecule which enables to monitor protein dynamics in solution or formation of non-covalent protein-ligand and protein-protein complexes. There are various types of chemical methods that are used for structural characterization of a polypeptide backbone. The most common ones include hydrogen/deuterium exchange (HDX) or non-reversible introduction of chemical tags, such as radicals modifying amino acid side chains.[1]

Hydrogen/deuterium exchange relies on the principle that the protein molecule in solution undergoes exchange of protein hydrogens with the hydrogens from surrounding heavy water.[2] Currently, stable deuterium oxide is used and deuterium incorporation into peptides is analyzed by nuclear magnetic resonance spectroscopy[3] or mass spectrometry (HDX-MS).[4] From the different groups of hydrogens present in the proteins, only those on backbone amides show exchanges detectable at a measurable time scale. However, this type of protein labeling is unstable due to its reversibility and this undesirable feature of the labeling reaction imposes strict requirements on the subsequent sample analysis and represents a risk that all labels could be lost before and during the protein analysis. Complementary to the inherently reversible HDX-MS is stable covalent labeling which provides the information on solvent accessibility and reactivity of amino acid side chains by means of an irreversible chemical modification. In contrast to HDX-MS, these modifications are permanent and resistant to sample workup.

Stable covalent labeling in combination with mass spectrometry was used to demonstrate its potential for structural biology quantifying the extent of lysine and arginine modifications on native and denatured lysozyme using acetic acid anhydride and 1,2-cyclohexadione.[5] Subsequently, the previous work was extended by the inclusion of additional model proteins/protein complexes and a variety of chemical probes targeting different amino acids. Indeed, the observed modifications accurately correlated with protein surface accessible area and established the stable protein covalent labeling as a suitable technique for structural characterization of proteins.[6–9] Although the stable covalent labeling provides enough information for protein structure dynamic studies, its spatial resolution is limited to several amino acid side chains – lysine, arginine and tyrosine; recently it was expanded to aspartic and glutamic acids as well.[10,11] This limitation has been overcome by discovery of hydroxyl radical footprinting that enables stable modification of the majority of amino acids found within the protein sequence.[12–14] In most cases, a good correlation was achieved between the degree of amino acids hydroxylation/oxidation and solvent accessibility. The value of the technique for probing protein folding and unfolding has been demonstrated on the example of gelsolin, an

actin binding protein.[15,16] The only drawback of this methodology is the unusual experimental instrumentation generating hydroxyl radicals where X-ray synchrotron irradiation is utilized to dissociate water molecules. To make the technique accessible to a broader range of laboratories, an alternative experimental design was introduced. Here, reactive oxygen species are generated by redox decomposition of hydrogen peroxide using Fe^{2+} ions (Fenton chemistry).[17] Since the Fenton reaction requires a high concentration of iron salt and EDTA as a complexing ligand, it raises a concern about the conservation of native protein structure throughout the course of labeling reaction, reproducibility of the results, and ease of sample preparation for mass spectrometric analysis. Therefore, rapid ultra-violet laser irradiation, also known as fast photo-oxidation of proteins (FPOP), has currently become a preferred method for studies of higher order structure of proteins and protein complexes.[18–20] The success of FPOP led to development of complementary methods based on iodine radicals and triplet carbenes (=carbene diradicals) as well.[21–23]

Even though FPOP is a promising technique for mapping protein solvent accessible area and it provides sufficient spatial resolution, there is still high demand for an alternative, less instrumentally complicated and inexpensive approach of protein stable covalent labeling. The discovery of stable, cyclic hypervalent iodine-fluoroalkyl reagents (Figure 1), also known as Togni reagents[24–26] and their ability to fluoroalkylate aromatic amino acids by ascorbic acid induced fluoroalkylation[27] suggests to capitalize on their high reactivity and use them for protein radical footprinting too.

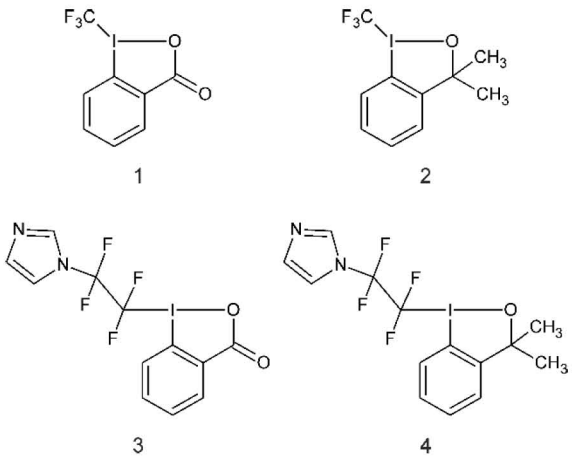


Figure 1. Structures of cyclic hypervalent iodine-fluoroalkyl reagents used in this study 1) acid type Togni-CF_3 reagent 2) alcohol type Togni-CF_3 reagent 3) acid type Togni-CF_2F_2Im reagent and 4) alcohol type Togni-CF_2CF_2Im reagent

Here, the performance of quench flow capillary set-up is presented for Fast FluoroAlkylation of Proteins (FFAP) at a time scale of 3 seconds. The instrumental setup consists of three syringe pumps and two solvent mixing T-pieces. The protein sample is premixed with Togni reagents: 1-(trifluoromethyl)-1λ^3-benzo[d][1,2]iodaoxol-3(1H)-one (Togni reagent II, acid type Togni-CF_3 reagent, **1**); 3,3-dimethyl-1-(trifluoromethyl)-1,3-dihydro-1λ^3-benzo[d][1,2]iodaoxole (alcohol type Togni-CF_3 reagent, Togni reagent I, **2**); 1-(1,1,2,2-tetrafluoro-2-(1H-imidazol-1-yl)ethyl)-1λ^3-benzo[d][1,2]iodaoxol-3(1H)-one (acid type Togni-CF_2F_2Im reagent, **3**) or 1-(2-(3,3-dimethyl-1λ^3-benzo[d][1,2]iodaoxol-1(3H)-yl)-1,1,2,2-tetrafluoroethyl)-1H-imidazole (alcohol type Togni-CF_2CF_2Im reagent, **4**) and the corresponding fluoroalkyl radicals are generated subsequently by introduction of ascorbic acid solution to the mixing chamber. The labeling pulse is quenched by addition of L-tryptophan solution.

The robustness of FFAP was tested by mapping structural differences of native horse heart apomyoglobin (aMb) versus holomyoglobin (hMb, complex with heme). The duration of labeling pulses was kept at three seconds. Nineteen and sixteen amino acid side chains out of 22 available aromatic amino acid residues were labeled when using either **1** or **2** and also **3** or **4**, respectively. The extent of trifluoromethylation was higher for 9 residues in aMb (compared to hMb) and for 3 residues in hMb (compared to aMb). Three amino acid side chains were only modified at aMb and any fold change was not observed for four residues. Likewise, the degree of imidazoletetrafluorethylation was higher for 6 residues in aMb (compared to hMb) and for 3 residues hMb (compared to aMb). Five amino acid side chains were only modified at aMb and two residues showed an opposite reactivity trend for **3/4** pair. Only F46, H48, F123 and F46, H48, H82, Y103, H113, F123 were not modified using **1/2** and **3/4** reagent, respectively. Since the solvent accessible area of aMb and hMb forms correlates well with previously published data using HDX[28],FPOP[29] and laser-initiated radical footprinting[30], we went on to test the applicability of FFAP to large protein-protein complexes. We elected to characterize the interaction interface of Haptoglobin (Hpt)/Hemoglobin (Hb) complex. Hpt and Hb were labeled with fluoroalkyl radicals separately and in complex. Of the Togni reagents, only the more reactive acid type derivatives were utilized, i.e., reagents **1** and **3**. In theory, both alpha and beta Hpt subunits contain 39 residues sensitive to modification, whereas Hbα and Hbβ contain 22 each. When using **1**, 35 out of 93 residues were not reactive under any conditions, and eighteen amino acids had an identical degree of modification, irrespective whether free or bound (Hbα – F44 and H88, Hbβ – H3, W16, F46, F78, F123, Y131, H144, Hptα – H23 and Hptβ – F205, Y242, Y272, H337, F339, H362, Y386, W398), eight residues (Hbα – F34 and F37, Hbβ – Y146, Hptα – H29 and Hptβ – F292, H295, Y298 and Y352) were selectively modified in the free state only, Hbα – H51 and Hptβ – Y370 were exclusively trifluoromethylated in complex. An increase of reactivity was observed for Hbβ – F43, Hptα – Y102 and Hptβ – W196, Y224 in complex, and finally seventeen amino side chains (Hbα –W15, Y25, H46, H90, Hbβ – Y36, W38, F42, F72, H98, F104, H147, Hptα – W133 and Hptβ – H166, H208, Y280, W284) were less modified in complexed state. Similar results were achieved using **3**. Since the modification plot was created for each reactive residue within Hb and Hpt sequences, it enabled correlation of our footprinting results to the Hb-Hpt structural model previously characterized using X-Ray diffraction spectroscopy (PDB: 4WJG).[31] Indeed, residues unreactive and less reactive in bound state were found to lie on the interaction interface of the Hb-Hpt complex which demonstrates the suitability of FFAP approach for the structural characterization of large protein complexes as well.

Methods

Preparation of labeling buffer

Ammonium bicarbonate was dissolved in LC-MS quality water to a final concentration of 50 mM and degassed by purging with helium. The pH was adjusted to 7.4 by bubbling using carbon dioxide.

Protein preparation

Lyophilized proteins (human 1,1-haptoglobin, horse heart myoglobin, human α,β-hemoglobin) were dissolved in the labeling buffer and purified using gel chromatography column (ENrichTM SEC 70, Biorad, USA). The final concentration of proteins was determined by UV VIS spectrophotometer (UV-VIS spectrophotometer DeNovix DS-11, DeNovix, USA). For preparation of apomyoglobin, 12 mg of holomyoglobin were dissolved in 5 ml of the labeling buffer. To the precooled solution of holomyoglobin, ice cold hydrochloric acid was dropwise added with external ice cooling to reach the final pH of 1.5. Then, 20 ml of butan-2-one (4 times the volume of labeling buffer) was added to the sample, transferred to the separatory funnel and shaken for 10 seconds to extract off the liberated heme. The aqueous, heme-free phase was collected and concentrated using Amicon® Ultra-15 centrifuge tubes with 10 kDa cut-off. Finally, the protein was purified by gel chromatography column (ENrichTM SEC 70, Biorad, USA). The concentration of both apo- and holomyoglobin was adjusted to 0.66 mg/ml.

The haptoglobin-hemoglobin complex was prepared by mixing 22 µl of α,β-hemoglobin at concentration of 0.46 mg/ml with 15 µl of 1,1-haptoglobin at concentration of 0.66 mg/ml. Hemoglobin solution was prepared by mixing 22 µl of α,β-hemoglobin at a concentration of 0.46 mg/ml with 15 µl of the labeling buffer. 1,1-haptoglobin solution was prepared by mixing 15 µl of 1,1 haptoglobin at concentration of 0.66 mg/ml with 22 µl of the labeling buffer. The amount of protein was 10 µg per reaction.

Radical labeling with Togni reagents

All four Togni reagents (**1**,**2**,**3**,**4**) were dissolved in dimethylsulfoxide to final concentration of 50 mg/ml. Thereafter, a Togni reagent (100 equiv) was added to the protein samples to final concentration of 7.5 mM . L-Ascorbic acid was dissolved in the labeling buffer to a final concentration of 0.4 mg/ml. Hamilton syringes were used to deliver the mixture of protein with Togni reagent and ascorbic acid (75 equiv) as a radical inducer to the mixing chamber according to the setup shown in **Figure 2**. L-Tryptophan solution (10 mg/ml) used as a quencher was delivered by the third Hamilton syringe to the quenching chamber. Flow rate, length and diameter of the capillaries were set to give a labeling time of three seconds. The samples were collected for 3 minutes and the final amount of labeled protein was 10 ng.

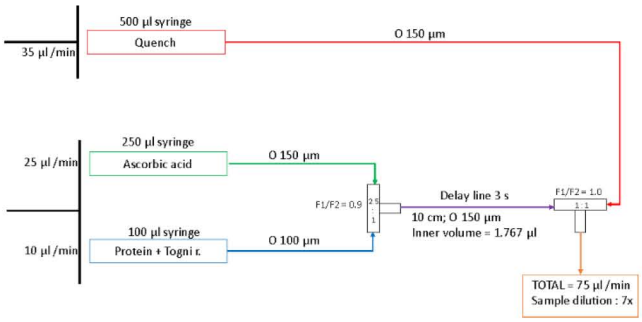


Figure 2. Scheme of FFAP labeling platform

Sample preparation for MS analysis

Samples were 2 times diluted with 100mM 4-ethylmorpholine buffer, ACN (90:10 v/v)), pH 8.5. α,β-hemoglobin and 1,1 haptoglobin were reduced with 10 mM Tris(2-carboxyethyl)phosphine (TCEP) at 56 °C for 10 minutes. Free cysteines were subsequently alkylated with 20 mM iodoacetamide (IAA) at room temperature for 20 minutes in the dark. 1,1-haptoglobin was deglycosylated overnight with PNgase F (protein:enzyme ratio 1:20). All samples were digested by trypsin/Lys-C (protein:enzyme ration 1:20) for two hours with addition trypsin/Lys-C digestion for another two hours. Digestion was stopped by addition of TFA to a final concentration of 0.1 %. Samples were dried by SpeedVac concentrator (Eppendorf).

LC-MS and LC-MS/MS analysis of labeled samples

LC-MS analysis was performed as described previously. Briefly, peptide mixture was reconstituted with 30 µl of 0.1% formic acid and 5 µl was injected using autosampler Agilent 1290 series UPLC system on desalting pre-column Luna® Omega Polar C18 5 µm, 100 Å, 0.3x30 mm. After 5 min desalting at a flow rate 20 µl/min, the peptides were separated by analytical column Luna® Omega Polar C18, 3 µm, 100 Å, 0.3x150 mm heated to 50 °C at a flow rate of 10 µl/min with acetonitrile gradient from 5% to 35% in 35 minutes. The chromatographic system was directly coupled with electrospray ionization source of solariX XR FT-ICR mass spectrometer (Bruker Daltonics, Germany) equipped with 15T superconducting magnet. The eluted peptides were analysed in a positive mode with one million transient data points over the range 250-2500 m/z. Final spectrum was created by averaging four following spectra with accumulation ions in collision cell for 0.2 s. The mass spectrometer was operated in data independent mode. Analyte at m/z 922.0098 (ESI-TOF tuning mix, Agilent Technologies, USA) was used as a lock mass. Fragmentation data was acquired with 15 eV collision voltage, 0.2 s MS/MS ion accumulation and 2 scans were accumulated per spectra.

Identification of modified residues

The raw data was processed by software Data Analysis 5.0 (Bruker Daltonics, Germany) and exported to the mascot generic files (mgf). The mfg files were uploaded into Mascot (Matrix Science, Great Britain)[32] search engine with following search setup: Protease – Trypsin/Lys-C with three missed cleavage, variable modification – oxidation of methionine, fluoralkylation of tryptophan, tyrosine, phenylalanine, histidine and cysteine, fixed modification – carbamidation of cysteines. Error for precursor ion was set at 5 ppm and for fragments ions at 0.05 Da. A mass shift for the modification by trifluormethyl radicals was 66.9784 Da and for imidazole-tetrafluorethyl radicals 166.0154 Da. All modifications were checked manually.

Quantification of modified residues

Ion intensities of the first isotope of the most intensive charge state were used for quantification. For quantification the same equation (**eq.1**) was used as for the FPOP experiment which was done by group of Michael Gross[23].

$$\textbf{eq.1} \qquad Extent\ of\ modification = \frac{\Sigma I_{mod}}{\Sigma(I_{mod} + I)}$$

I_{mod} denotes the intensity of the peptide with modified residue and I the intensity of the peptides with non-modified residue. As a statistical tool standard deviation, P-value and T-test were used. Statistically relevant results were plotted into X-ray structure and visualized by software PyMol 1.7.2.1, Schrödinger, USA.

Results and discussion

Apo- and holomyglobin were chosen as a model proteins for radical labeling with Togni reagents in presence of ascorbic acid and also for radical labeling by fluoralkyl radicals generated [30], because the three dimensional structure was solved for both its forms by X-ray crystallography[33], HDX[28] and FPOP[29] For the FFAP, the data were obtained by quantification of all modified residues in holomyoglobin and apomyoglobin for all 4 reagents. Acid type Togni-CF_3 reagent **1** yielded the highest level of chemical modification: 4.3 % in the holomyoglobin and 5.6 % in the apomyoglobin per aromatic residue. Slightly lower level of modification of aromatic residues - 4.1 % for holomyoglobin and 4.8 % for apomyoglobin was observed for the alcohol type Togni-CF_3 reagent **2** as the source of trifluormethyl radical. For the acid type Togni-CF_2CF_2Im reagent **3**, only 1.5 % of holomyoglobin and 3.0 % of apomyoglobin per aromatic residue were modified. The imidazol-tetrafluorethyl radicals, generated from alcohol-type Togni-CF_2CF_2Im reagent **4**, modified the aromatic residues at 2.0 % in holomyoglobin and 1.2 % in apomyoglobin.

In total, 19 out of 22 aromatic acids were modified in apo- or holomyoglobin, when **1** was used (**Figure 2A**). Statistically significant changes between apo- and holomyoglobin were observed in 14 aromatic residues (W7, W14, H24, F33, H36, F43, H64, H82, H93, H97, H113, H119, F138, Y146, F151). For aromatic amino acids W7, W14, H24, F33, H36, F43, H64, H97, F138, Y146 and F151, a higher level of modification was observed for apomyoglobin compared to holomyoglobin. On the other hand, the aromatic amino acids H82, H113 and H119 were modified more efficiently in the holo-form compared to the apo-form. This observation is consistent with the localization of these 3 hydrophilic histidine residues on the surface of the holomyoglobin and their better solvent accessibility in this region compared to apomyoglobin[1]. Furthermore, H119 is located on the beginning of an alpha helix with outside orientation, near a flexible loop of the protein chain. On the other hand, amino acids W7, W14 and H24 were less modified in the holomyoglobin, most likely as a result of the higher degree of protein packing and resulting diminished reactivity compared to apomyoglobin. The other amino acids, which are less modified in holomyoglobin are located in the heme binding site. Conversely, they are more solvent accessible in apomyoglobin because of heme absence.

Very similar results were achieved when alcohol type Togni-CF_3 reagent **2**, another source of trifluoromethyl radical, was used. Here, 19 out of 22 aromatic residues were labeled and 14 were statistically relevant (**Figure 2B**). The amino acids H81, H82 and H119, located on the surface, revealed increased modification in holomyoglobin in comparison with apomyoglobin, whereas amino acids F33, H36, F43, H64, H93, H97, F138, Y146 and F151 situated in the heme binding site of the myoglobin were less modified.

[1] It has been described that removal of heme from holomyoglobin induces a global conformational change that leads to a less rigid structure with higher structural flexibility, particularly in the heme-binding domain.

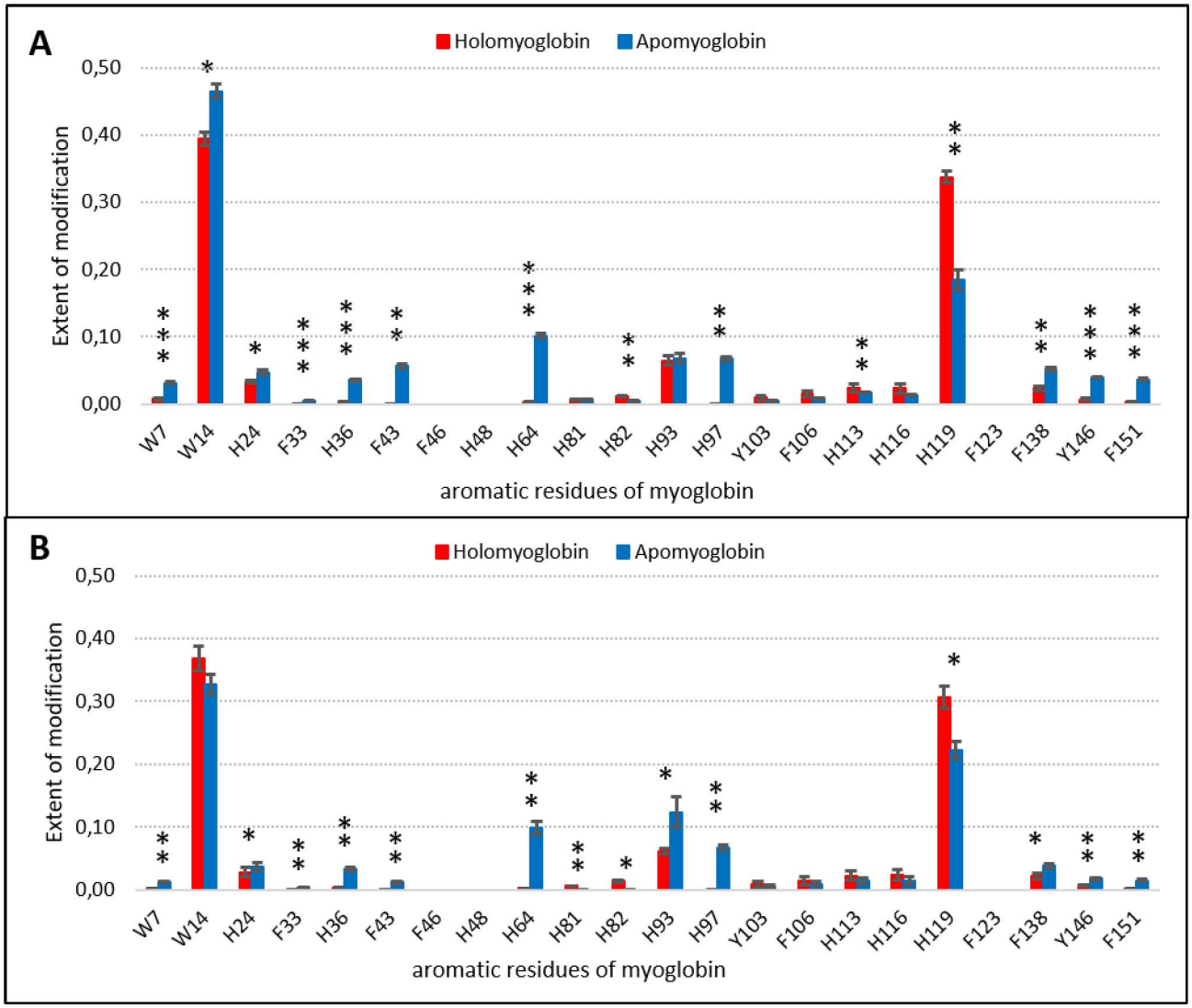


FIGURE 2: Quantification of aromatic residues of apo- and holomyoglobin for acid type Togni-CF_3 reagent **1** (A) and for alcohol type Togni-CF_3 reagent **2** (B). Red bars represent the extent of holomyoglobin modification and blue bars represent apomyoglobin modification, P-value (*** - P < 0,005; ** - P < 0,01; * - P < 0,05).

The aromatic residues W7 and H24, oriented towards the core of the protein, were also less modified in holomyoglobin. The most likely explanation is the increased degree of rigification in heme binding domains. The modification by imidazole-tetrafluorethyl radicals formed by reductive decomposition Togni-CF_2CF_2Im reagent **3** provided slightly different results. In total, 16 out of 22 aromatics residues underwent modification (W7, W14, H24, F33, H36, F43, H64, H81, H93, H97, F106, H116, H119, F138, Y146, F151) and all showed statistically significant difference between apo- and holomyglobin. The increased modification of aromatic amino acids F106, H116 and H119 in holomyoglobin is most likely caused by their localization at the single alpha helix with orientation towards the solvent (**Figure 3A**). The other modified amino acids F33, H36, F43, H64, H81, H93, H97, F138, Y146 and F151a were modified more intensively in apomyoglobin. The three amino acids W7, W14 and H24, which are oriented towards the core of the protein, were modified with higher intensity in apomyoglobin compared to holomyoglobin.

The fourth Togni reagent, the alcohol-type Togni-CF_2CF_2Im reagent **4**, modified 15 out of 22 aromatic residues with more significant variation between apo- and holomyoglobin. Like in the case of acid type Togni-CF_2CF_2Im reagent **3**, the aromatic residues F106, H113 and H119 were modified in holomyoglobin more efficiently than in apomyglobin. In contrast to **3**, the residues W14 and H24 were modified more effectively in holomyoglobin, when **4** was used. The other modified aromatic residues are located in the active site of myoglobin and are less modified in holomyoglobin than in apomyoglobin.

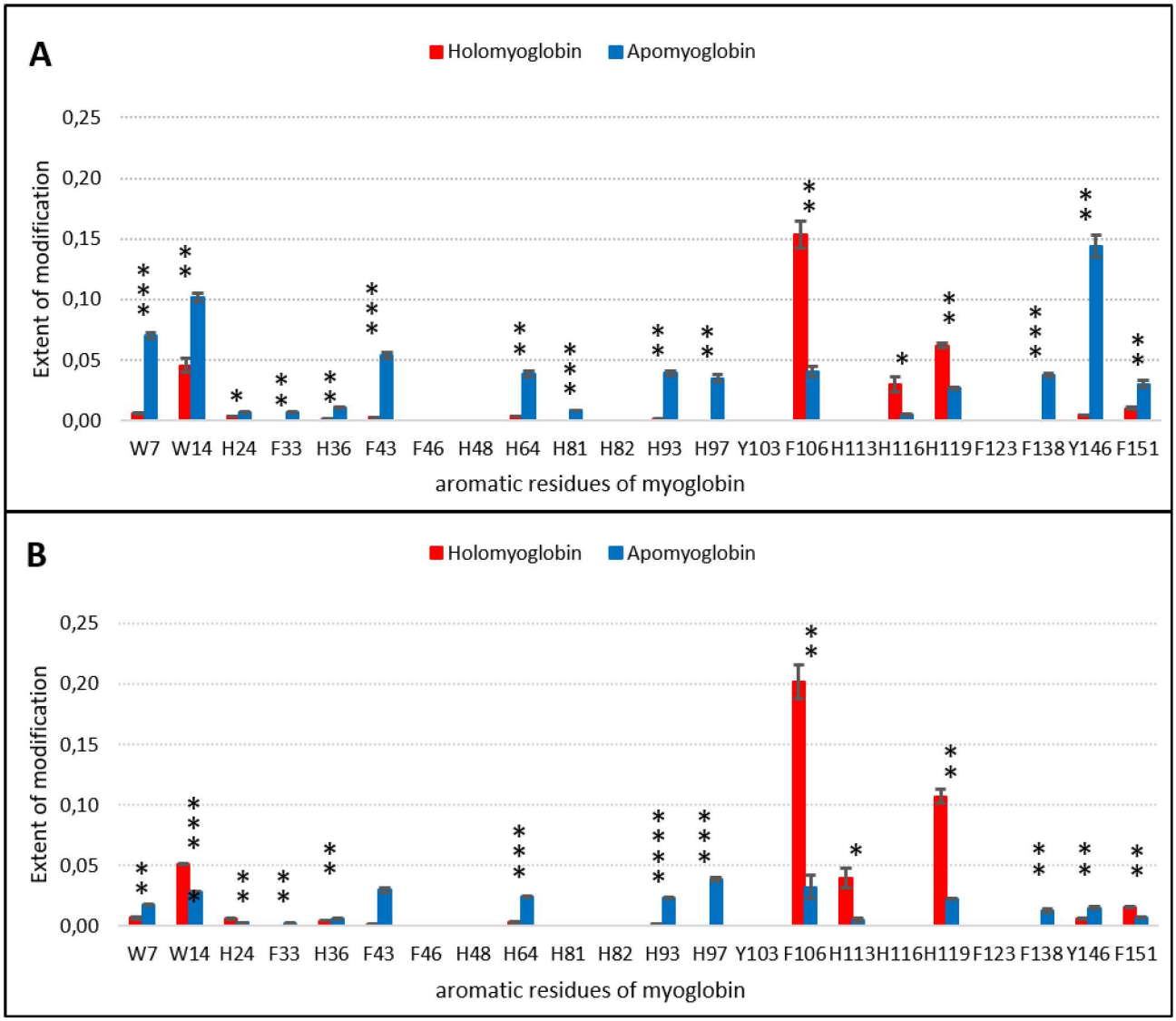


FIGURE 3: Quantification of aromatic residues of apo- and holomyoglobin for acid type Togni-CF_2CF_2Im reagent **3** (A) and for alcohol type Togni-CF_2CF_2Im reagent **4** (B). Red bars represent the extent of modification for holomyoglobin and blue bars for apomyoglobin, P-value (*** - $P < 0{,}005$; ** - $P < 0{,}01$; * - $P < 0{,}05$).

To sum up, experiments with all four Togni reagents provided similar structural information about two regions in which major differences between apomyoglobin and holomyglobin were observed (**Figure 4**). The most affected region is the active site (= heme binding site) of the protein and the A helix (Figure 4 – Red), where all aromatic acids (F33, H36, F43, H64, H81, H93, H97, F138, Y146, F151) are less modified in the presence of heme. The other region (Figure 4 – Blue), composed of two helixes G and H connected by a small linker, is more effectively modified in holomyoglobin. The CD spectroscopy has shown that in the absence of heme in the protein structure of myoglobin, helixes C, D and F are destabilized.[34] This corresponds to the observed increased modification of aromatic amino acids situated at these three helixes by fluoroalkyl radicals. The FPOP data published by Cheng et al[30] using the same system of apo- and holomyoglobin are in good accordance with our FFAP results. In their experiments, the helixes E, F, and H were modified more efficiently by hydroxyl radicals in the absence of heme. Using our FFAP method, modifications on helixes A and C were also observed. The helix C is a part of the active site of the myoglobin, whereas helix A is not located in the active site. Aromatic residues W7 and W14 of the A helix are oriented towards the core of the protein.

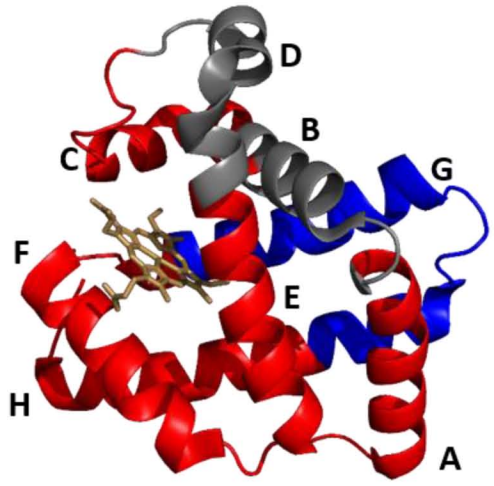


***FIGURE* 4:** Visualization of the regions with the highest degree of modification and the difference between apo- and holomyoglobin. The red regions are less modified in holomyoglobin and the blue regions are more modified in holomyoglobin.

A small difference in modification of aromatic residues was observed when comparing the two types of Togni reagents yielding the same type of radical. Togni reagents possessing the cyclic carboxylate group in the iodane (the so-called acid type Togni reagents) typically yield higher degrees of fluoroalkylation than alcohol type Togni reagents (**Supplemental Figure S4**). For further experiments with 1,1-haptoglobin/α,β-hemoglobin complex, only the acid type Togni-CF_3 reagent **1** was employed. The total level of modification by **1** was 2.5 % for Hbα subunit and 6.4 % for Hbβ subunit in uncomplexed Hbαβ. When the Hbαβ was in a complex with Hpt, the level of modification for Hbα was 0.9 % and for Hbβ 4.4 % per aromatic residue. The level of modification of Hpt alone was 2.7 % and 1.8 % when Hpt was in complex with Hbαβ. The acid type Togni-CF_2CF_2Im **3** reagent forming imidazoletetrafluorethyl radicals modified the Hbα subunit up to 2.7 % and in Hbαβ/Hpt complex, up to 1.5% of the aromatic amino acids of Hbα subunit underwent modification(**Supplemental Figure S7**).[29] The Hbβ was modified up to 5.2 % and in Hbβ/Hpt complex, up to 4.2 % per aromatic residue of Hbβ were modified. Haptoglobin was modified up to 0.9 % per aromatic residue in both free and complexed state. Ten out of 21 aromatic residues in Hbα were modified by **1**. A significant difference in modification between the Hbα subunit in uncomplexed Hbαβ and Hpt/Hbαβ complex was detected for 6 aromatic residues. The amino acids W15, Y25, F37, H46 and H90 were less modified in Hpt/Hbαβ, whereas amino acid H51 was modified more effectively in Hpt/Hbαβ complex (**Supplemental Figure S7**). The aromatic residues H90 and F37 are located at the interaction interface of the Hpt/Hbαβ complex and are therefore less modified. The other observed modifications were identified at different part of Hbα and their changes in the level of modifications are probably caused by dynamics of the molecule. The overall decrease of modification in Hpt/Hbαβ complex is manifested in stabilization of the Hbα unit. Sixteen out of 22 aromatic residues were modified in Hbβ. Statistically significant changes were obtained for 9 residues. The amino acids Y36, W38, F42, F72, H98, F104, Y146 and H1478 were less modified in the Hpt/Hbαβ complex. The F43 was modified at higher level in this complex. All these modified amino acids are located on the interaction interface of the Hpt/Hbαβ complex. Twenty-three out of 38 aromatic acids of Hpt were modified by trifluormethyl radicals. Three modified aromatic residues are located in the alpha subunit of haptoglobin. The amino acids H105 and W133 were modified less intensively in Hpt/Hbαβ complex, whereas amino acid Y102 was modified more intensively in complex. The rest of modified residues are located in the beta subunit of Hpt. Aromatic residues H166, H208, Y280, W284, F292, H295, Y298 and Y352 were less modified in the complex. The three amino acids W196, Y224 and Y370 were modified more intensively in the Hpt/Hbαβ complex. All significant modifications were visualized into the previously published X-Ray structure of the Hpt/Hbαβ complex (**Figure 5**). With the exception of H208, which is located on the flexible linker, all the less modified aromatic residues are positioned on the interaction interface with both subunits of Hb. The three amino acids, which are modified to higher levels in a complex, are located on the surface of the molecule, out of the interaction interface.

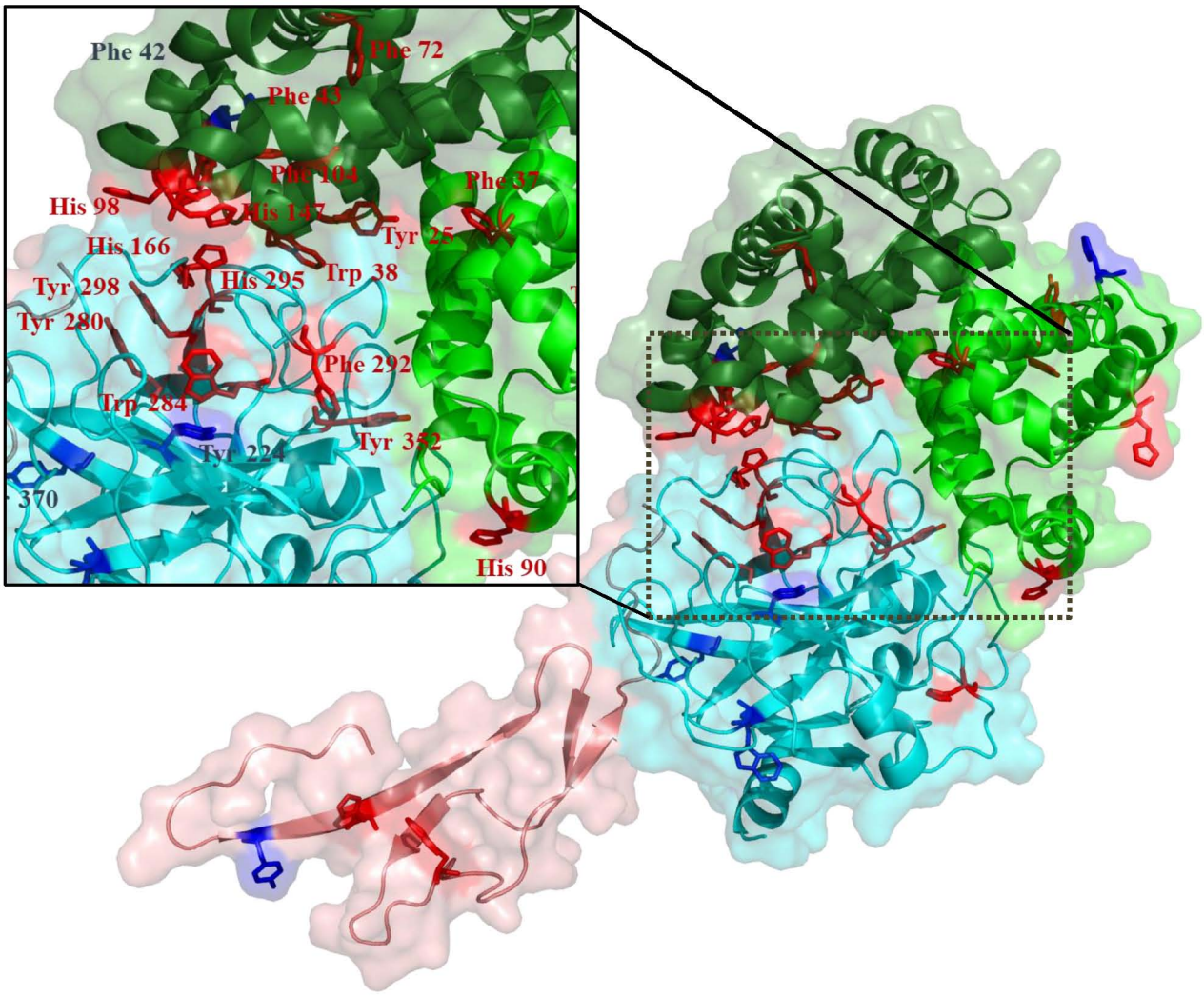


FIGURE 5: Structure of Hpt/Hbαβ complex with highlighted modified residues after reaction with **3**. Red residues are less modified in Hpt/Hbαβ complex and blue are modified to higher levels in Hpt/Hbαβ complex. Subunits of complex are highlighted as follows: alpha subunit of Hpt - pink, beta subunit of Hpt-light blue, alpha subunit of Hb - dark green and beta subunit of Hpt/Hbαβ - light green. Zoom – Interaction interface of Hpt/Hbαβ complex.

Conclusion

A new radical labeling technique based on fluoroalkylation of proteins by reductive decomposition of Togni reagents by sodium ascorbate was developed and optimized for modification of aromatic amino acids. The functionality of the method was tested using model systems of apo- and holomyoglobin and haptoglobin/hemoglobin complex. The results show a very high degree of correlation with previously published structural data and confirm the applicability of Togni reagent-mediated FFAP as an alternative method for covalent labeling in applications such as protein fingerprinting, epitope mapping and surface mapping of proteins (and their complexes) in general.

ASSOCIATED CONTENT

Supporting Information. Mass spectra for all reactions of myoglobin with Togni reagents **1-4**. Extracted ion chromatograms for peptide 1-16 derived from myoglobin. Structures of myoglobin modified by all Togni reagent. Quantification plots for haptoglobin-hemoglobin complex for **1** and **3**. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Zdeněk Kukačka and Petr Novak, Institute of Microbiology of the Czech Academy of Sciences, BIOCEV, Prumyslova 595, 252 50 Vestec, Czech Republic, zdenek.kukacka@biomed.cas.cz and pnovak@biomed.cas.cz.

Conflict of interest

The company CF Plus Chemicals commercializes the iodanes and the fluoroalkylation technology (a patent application EP19172756) appearing in this study.

Present Addresses

†If an author's address is different than the one given in the affiliation line, this information may be included here.

Author Contributions

The manuscript was conceived from contributions of all authors.

Funding Sources

The Czech Science Foundation (grant number 19-16084S), the Ministry of Education of the Czech Republic (program "NPU II" project LQ1604), and European Commission H2020 (EPIC-XS - grant agreement ID: 823839 and EU FT-ICR MS - grant agreement ID: 731077).

ACKNOWLEDGMENT

This work was mainly supported by the Czech Science Foundation (grant number 19-16084S), the Ministry of Education of the Czech Republic (program "NPU II" project LQ1604), European Commission H2020 (EPIC-XS - grant agreement ID: 823839 and EU FT-ICR MS - grant agreement ID: 731077), and, in part, by the Czech Academy of Sciences (RVO61388971). We acknowledge Structural mass spectrometry core facility of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2018127).

ABBREVIATIONS

FFAP, fast fluoroalkylation of proteins; FPOP, fast photo-oxidation of proteins; IAA, iodacetamide; HDX, hydrogen/deuterium exchange; MS, mass spectrometry; TCEP, tris(2-carboxyethyl)phosphine, TFA, trifluoracetic; acid type Togni-CF_3 reagent, 1-(trifluoromethyl)-1λ^3-benzo[d][1,2]iodaoxol-3(1H)-one; alcohol type Togni-CF_3 reagent, 3,3-dimethyl-1-(trifluoromethyl)-1,3-dihydro-1λ^3-benzo[d][1,2]iodaoxole; acid type Togni-CF_2F_2Im reagent, 1-(1,1,2,2-tetrafluoro-2-(1H-imidazol-1-yl)ethyl)-1λ^3-benzo[d][1,2]iodaoxol-3(1H)-one; alcohol type Togni-CF_2CF_2Im reagent, 1-(2-(3,3-dimethyl-1λ^3-benzo[d][1,2]iodaoxol-1(3H)-yl)-1,1,2,2-tetrafluoroethyl)-1H-imidazole

REFERENCES

(1) Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, acs.chemrev.9b00815. https://doi.org/10.1021/acs.chemrev.9b00815.

(2) Benson, E. E.; Linderstrom-Lang, K. Deuterium Exchange between Myoglobin and Water. *Biochim. Biophys. Acta* **1959**, *32*, 579–581. https://doi.org/10.1016/0006-3002(59)90649-3.

(3) Englander, S. W.; Kallenbach, N. R. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Q. Rev. Biophys.* **1983**, *16* (4), 521–655. https://doi.org/10.1017/S0033583500005217.

(4) Zhang, Z.; Smith, D. L. Determination of Amide Hydrogen Exchange by Mass Spectrometry: A New Tool for Protein Structure Elucidation. *Protein Sci.* **1993**, *2* (4), 522–531. https://doi.org/10.1002/pro.5560020404.

(5) Suckau, D.; Mak, M.; Przybylski, M. Protein Surface Topology-Probing by Selective Chemical Modification and Mass Spectrometric Peptide Mapping. *Proc. Natl. Acad. Sci.* **1992**, *89* (12), 5630–5634. https://doi.org/10.1073/pnas.89.12.5630.

(6) Glocker, M. O.; Borchers, C.; Fiedler, W.; Suckau, D.; Przybylski, M. Molecular Characterization of Surface Topology in Protein Tertiary Structures by Amino-Acylation and Mass Spectrometric Peptide Mapping. *Bioconjug. Chem.* **1994**, *5*, 583–590.

(7) Fiedler, W.; Borchers, C.; Macht, M.; Deininger, S. O.; Przybylski, M. Molecular Characterization of a Conformational Epitope of Hen Egg White Lysozyme by Differential Chemical Modification of Immune Complexes and Mass Spectrometric Peptide Mapping. *Bioconjug. Chem.* **1998**, *9* (2), 236–241. https://doi.org/10.1021/bc970148g.

(8) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. A Top-down Method for the Determination of Residue-Specific Solvent Accessibility in Proteins. *J. Mass Spectrom.* **2004**, *39* (3), 322–328. https://doi.org/10.1002/jms.587.

(9) Antonovic, L.; Hodek, P.; Smrcek, S.; Novák, P.; Sulc, M.; Strobel, H. W. Heterobifunctional Photoaffinity Probes for Cytochrome P450 2B. *Arch. Biochem. Biophys.* **1999**, *370* (2), 208–215. https://doi.org/10.1006/abbi.1999.1408.

(10) Zhang, H.; Shen, W.; Rempel, D.; Monsey, J.; Vidavsky, I.; Gross, M. L.; Bose, R. Carboxyl-Group Footprinting Maps the Dimerization Interface and Phosphorylation-Induced Conformational Changes of a Membrane-Associated Tyrosine Kinase. *Mol. Cell. Proteomics* **2011**, *10* (6), M110.005678. https://doi.org/10.1074/mcp.m110.005678.

(11) Kaur, P.; Tomechko, S. E.; Kiselar, J.; Shi, W.; Deperalta, G.; Wecksler, A. T.; Gokulrangan, G.; Ling, V.; Chance, M. R. Characterizing Monoclonal Antibody Structure by Carboxyl Group Footprinting. *MAbs* **2015**, *7* (3), 540–552. https://doi.org/10.1080/19420862.2015.1023683.

(12) Maleknia, S. D.; Chance, M. R.; Downard, K. M. Electrospray-Assisted Modification of Proteins: A Radical Probe of Protein Structure. *Rapid Commun. Mass Spectrom.* **1999**, *13* (23), 2352–2358. https://doi.org/10.1002/(SICI)1097-0231(19991215)13:23<2352::AID-RCM798>3.0.CO;2-X.

(13) Maleknia, S. D.; Ralston, C. Y.; Brenowitz, M. D.; Downard, K. M.; Chance, M. R. Determination of Macromolecular Folding and Structure by Synchrotron X-Ray Radiolysis Techniques. *Anal. Biochem.* **2001**, *289* (2), 103–115. https://doi.org/10.1006/abio.2000.4910.

(14) Chance, M. R. Unfolding of Apomyoglobin Examined by Synchrotron Footprinting. *Biochem. Biophys. Res. Commun.* **2001**, *287* (3), 614–621. https://doi.org/10.1006/bbrc.2001.5628.

(15) Kiselar, J. G.; Janmey, P. A.; Almo, S. C.; Chance, M. R. Visualizing the Ca2+-Dependent Activation of Gelsolin by Using Synchrotron Footprinting. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (7), 3942–3947. https://doi.org/10.1073/pnas.0736004100.

(16) Kiselar, J. G.; Janmey, P. A.; Almo, S. C.; Chance, M. R. Structural Analysis of Gelsolin Using Synchrotron Protein Footprinting. *Mol. Cell. Proteomics* **2003**, *2* (10), 1120–1132. https://doi.org/10.1074/mcp.M300068-MCP200.

(17) Sharp, J. S.; Becker, J. M.; Hettich, R. L. Protein Surface Mapping by Chemical Oxidation: Structural Analysis by Mass Spectrometry. *Anal. Biochem.* **2003**, *313* (2), 216–225. https://doi.org/10.1016/S0003-2697(02)00612-7.

(18) Sharp, J. S.; Becker, J. M.; Hettich, R. L. Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry. *Anal. Chem.* **2004**, *76* (3), 672–683. https://doi.org/10.1021/ac0302004.

(19) Aye, T. T.; Low, T. Y.; Sze, S. K. Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* **2005**, *77* (18), 5814–5822. https://doi.org/10.1021/ac050353m.

(20) Hambly, D. M.; Gross, M. L. Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2057–2063. https://doi.org/10.1016/j.jasms.2005.09.008.

(21) Chen, J.; Cui, W.; Giblin, D.; Gross, M. L. New Protein Footprinting: Fast Photochemical Iodination Combined with Top-down and Bottom-up Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (8), 1306–1318. https://doi.org/10.1007/s13361-012-0403-1.

(22) Zhang, B.; Rempel, D. L.; Gross, M. L. Protein Footprinting by Carbenes on a Fast Photochemical Oxidation of Proteins (FPOP) Platform. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (3), 552–555. https://doi.org/10.1007/s13361-015-1313-9.

(23) Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chemie Int. Ed.* **2017**, *56* (45), 14007–14010. https://doi.org/10.1002/anie.201706697.

(24) Eisenberger, P.; Gischig, S.; Togni, A. Novel 10-I-3 Hypervalent Iodine-Based Compounds for Electrophilic Trifluoromethylation. *Chem. - A Eur. J.* **2006**, *12* (9), 2579–2586. https://doi.org/10.1002/chem.200501052.

(25) Matoušek, V.; Pietrasiak, E.; Schwenk, R.; Togni, A. One-Pot Synthesis of Hypervalent Iodine Reagents for Electrophilic Trifluoromethylation. *J. Org. Chem.* **2013**, *78* (13), 6763–6768. https://doi.org/10.1021/jo400774u.

(26) Matoušek, V.; Václavík, J.; Hájek, P.; Charpentier, J.; Blastik, Z. E.; Pietrasiak, E.; Budinská, A.; Togni, A.; Beier, P. Expanding the Scope of Hypervalent Iodine Reagents for Perfluoroalkylation: From Trifluoromethyl to Functionalized Perfluoroethyl. *Chem. - A Eur. J.* **2016**, *22* (1), 417–424. https://doi.org/10.1002/chem.201503531.

(27) Rahimidashaghoul, K.; Klimánková, I.; Hubálek, M.; Korecký, M.; Chvojka, M.; Pokorný, D.; Matoušek, V.; Fojtík, L.; Kavan, D.; Kukačka, Z.; Novák, P.; Beier, P. Reductant-Induced Free Radical Fluoroalkylation of Nitrogen Heterocycles and Innate Aromatic Amino Acid Residues in Peptides and Proteins. *Chem. – A Eur. J.* **2019**, *25* (69), 15779–15785. https://doi.org/10.1002/chem.201902944.

(28) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 KDa Protein. *J. Am. Chem. Soc.* **2009**, *131* (35), 12801–12808. https://doi.org/10.1021/ja904379w.

(29) Vahidi, S.; Stocks, B. B.; Liaghati-Mobarhan, Y.; Konermann, L. Mapping PH-Induced Protein Structural Changes Under Equilibrium Conditions by Pulsed Oxidative Labeling and Mass Spectrometry. *Anal. Chem.* **2012**, *84* (21), 9124–9130. https://doi.org/10.1021/ac302393g.

(30) Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting. *Angew. Chemie Int. Ed.* **2017**, *56* (45), 14007–14010. https://doi.org/10.1002/anie.201706697.

(31) Stødkilde, K.; Torvund-Jensen, M.; Moestrup, S. K.; Andersen, C. B. F. Structural Basis for Trypanosomal Haem Acquisition and Susceptibility to the Host Innate Immune System. *Nat. Commun.* **2014**, *5* (1), 5487. https://doi.org/10.1038/ncomms6487.

(32) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(33) Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; Mauk, A. G.; Smith, M.; Brayer, G. D. A Myoglobin Variant with a Polar Substitution in a Conserved Hydrophobic Cluster in the Heme Binding Pocket. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1997**, *1341* (1), 1–13. https://doi.org/10.1016/S0167-4838(97)00064-2.

(34) Hughson, F.; Wright, P.; Baldwin, R. Structural Characterization of a Partly Folded Apomyoglobin Intermediate. *Science (80-.).* **1990**, *249* (4976), 1544–1548. https://doi.org/10.1126/science.2218495.

ARTICLE V

Fiala J., Kukačka Z., Novák P.

Influence of cross-linker polarity on selectivity towards lysine side chains.

J. Proteomics 2020, 218, 103716.

My contribution: Conception of experiments, conducting experiments (chemical cross-linking, LC-MS/MS analysis), data analysis and interpretation, manuscript writing

Journal of Proteomics 218 (2020) 103716

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Influence of cross-linker polarity on selectivity towards lysine side chains

Jan Fiala[a,b], Zdeněk Kukačka[a], Petr Novák[a,b,*]

[a] Institute of Microbiology of the Czech Academy of Sciences, Vestec, Czech Republic
[b] Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic

ARTICLE INFO

Keywords:
Structural mass spectrometry
Protein structure
Chemical cross-linking
Disuccinimidyl glutarate
Bis(sulfosuccinimidyl) glutarate
Bovine serum albumin

ABSTRACT

The combination of chemical cross-linking and mass spectrometry is currently a progressive technology for deriving structural information of proteins and protein complexes. In addition, chemical cross-linking is a powerful tool for stabilizing macromolecular complexes for single particle cryo-electron microscopy. Broad pallets of cross-linking chemistry, currently available for the majority of cross-linking experiments, still rely on the amine-reactive N-hydroxysuccinimide esters targeting mainly N-termini and lysine side chains. These cross-linkers are divided into two groups: water soluble and water insoluble; and research teams prefer one or another speculating on the benefits of their choice. However, the effect of cross-linker polarity on the outcome of cross-linking reaction has never been studied. Herein, we use both polar (bis(sulfosuccinimidyl) glutarate) and non-polar (disuccinimidyl glutarate) cross-linkers and systematically investigated the impact of cross-linker hydrophobicity on resulting distance constraints, using bovine serum albumin as a model protein.

Significance: Even though the amine reactive BS2G and DSG cross-linkers have the same length of spacer and are based on N-hydroxysuccinimidic group, our data showed that each of them formed preferentially different cross-links. We demonstrated that the choice of cross-linker can have a significant impact on the output data for structural characterization of biomolecules. Using equimolar mixtures of DSG with d6-BS2G, and BS2G with d6-DSG, we established that the polar BS2G preferentially bound to polar regions of modified molecule, whereas non-polar DSG bound to hydrophobic regions. This phenomenon established that the mixture of polar and non-polar cross-linkers acted as an efficient tool for the determination of distance constraints in proteins.

1. Introduction

Combination of chemical cross-linking and mass spectrometry (CXMS) is a popular technique for structural characterization of proteins and their complexes. Since the inception of CXMS's use for single protein structural study two decades ago [1], it is expanding as an alternative technique to high resolution spectroscopic methods of structural biology. The advantages of CXMS are small sample requirement, no limitation in protein size, native environment working condition, high sensitivity and fast data processing. All these aspects make CXMS very effective for proteins or their complexes that are difficult to solve by crystallography or NMR alone. In the majority of experiments reported, a protein or protein complex reacted with a bifunctional cross-linker that introduced a covalent bond between residues that were in close proximity [2]. The cross-linked protein was then digested, and the generated peptide mixture was analyzed by mass spectrometry for identifying the cross-linked peptides [3]. Such distance constraints were subsequently utilized for structural model design [4–6].

CXMS is well known for generating distance constraints for building the structural models of proteins and protein complexes [7]. In this regard, a diverse repertoire of the cross-linking chemistry is currently available utilizing homobifunctional [1,8,9], heterobifunctional [9,10] and photo-active [11,12] chemical probes. The number of restraints is insufficient to build a reliable structural model ab-initio, if one restraint is required per ten amino acids [4,5]. To increase the number of restraints, higher amount of cross-linking probe can be considered for the reaction. However, high concentration of cross-linker triggers an artificial conformation change due to the disruption of local electrostatic interactions. Further structural perturbations are caused by stabilizing a low-populated conformation or by limiting protein flexibility. Artificial conformational change exposes previously protected site and makes it accessible to the cross-linker [13]. Such findings advocate for a novel protocol allowing the detection of sufficient number of distance constraints by keeping the concentration of cross-linker low.

A broad pallet of cross-linking chemistry, currently available for the majority of cross-linking experiment still rely on the amine-reactive N-hydroxysuccinimide (NHS) esters targeting mainly N-termini and lysine side chains. Thus, our efforts focus on these chemical probes in detail.

* Corresponding author at: Institute of Microbiology of the Czech Academy of Sciences, Průmyslová 595, 252 50 Vestec, Czech Republic.
E-mail address: pnovak@biomed.cas.cz (P. Novák).

https://doi.org/10.1016/j.jprot.2020.103716
Received 6 December 2019; Received in revised form 2 February 2020; Accepted 19 February 2020
Available online 20 February 2020
1874-3919/

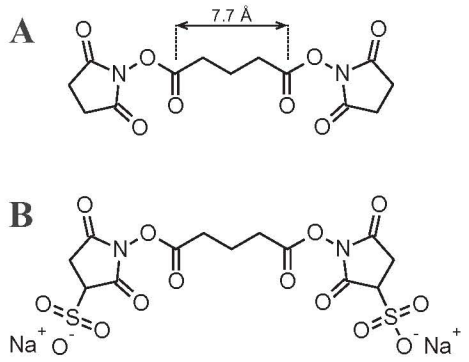


Fig. 1. Chemical structure of primary amine reactive N-hydroxysuccinimide ester (NHS ester) based cross-linkers used in this study. (A) Disuccinimidyl glutarate (DSG) and (B) bis(sulfosuccinimidyl)glutarate (BS2G). BS2G contains two additional sulfonate groups (-SO3) which give its negative charge and make it water-soluble.

These homobifunctional NHS-esters can be, in principle, divided into two categories: water soluble and water insoluble; and researchers preferred one or another based on their benefits. Although both hydrophilic [1,14] and hydrophobic [15,16] cross-linkers were successfully utilized to derive restraints for a design of structural model of proteins and protein complexes, the effect of cross-linker polarity on the outcome of the cross-linking reaction has not been studied yet. Herein, we used both polar (bis(sulfosuccinimidyl) glutarate, BS2G) and non-polar (disuccinimidyl glutarate, DSG) cross-linkers (Fig. 1) and systematically investigated the impact of cross-linker hydrophobicity on resulting distance constraints. Bovine serum albumin (BSA) was incubated with different mixtures consisting of regular and isotopically labeled probes, and the isotopic signature was exploited to determine the origin of identified cross-links. Subsequent quantification ratio of polar or non-polar probes in situations was done where both chemical probes form cross-links [17]. Such approach enabled the identification of five and two cross-links favoring DSG and BS2G, respectively, and 32 cross-links without significant preferences for cross-linker hydrophobicity. Based on these observations, BSA was also cross-linked using BS2G, DSG alone and with their equimolar mixture. Further, 46 and 47 cross-links were identified for BS2G and DSG, respectively, and 67 cross-links were determined for BS2G/DSG equimolar mixture. The ratio of cross-linker over the protein was kept identical to experiments where only polar or non-polar probe was taken into account.

2. Materials and methods

2.1. Chemicals and materials

BSA was purchased from Merck (USA). Disuccinimidyl glutarate (DSG), bis(sulfosuccinimidyl)glutarate (BS2G), and its six deuterated analogs were bought from Creative Molecules Inc. (USA). In addition, control experiments were prepared with DSG and BS2G cross-linkers and its four deuterated analogs obtained from ProteoChem (USA). Trypsin/Lys-C mass spectrometry grade proteases from Promega Corporation (USA) were used for digestion. All solvents used for HPLC analysis were LC/MS grade, purchased from Thermo Scientific (USA). Additional chemicals reported in this paper were purchased from Merck (USA) with the highest available purity.

2.2. Chemical cross-linking

DSG, BS2G and equimolar mixtures composed of DSG/BS2G, DSG/d6-BS2G and d6-DSG/BS2G were used for modification of 10 μM solution of BSA in 20 mM HEPES, 20 mM NaCl (pH 7.5) buffer. Additionally, control experiments using mixtures of DSG/d4-DSG and BS2G/d4-BS2G were performed. All previously mentioned cross-linkers and mixtures were prepared to a concentration of 10 mg/ml using anhydrous DMSO and added to BSA solution in 100× molar excess. Reactions were done for 30 min at room temperature. Before digestion, disulfide bonds were reduced with 20 mM Tris(2-carboxyethyl)phosphine (TCEP) at 56 °C for 20 min. Free cysteines were alkylated with 20 mM iodoacetamide (IAA) at room temperature for 20 min in the dark. After modification, samples were diluted by 5 times with a solution of 50 mM 4-ethylmorpholine buffer (pH 8.5):AcN (80:20 *v*/v); subsequently, trypsin/Lys-C was added (protein:enzyme ratio 1:20). Samples were digested overnight at 37 °C. Digested mixture was quenched with TFA to a final concentration of 0.1% and subsequently dried by SpeedVac (Eppendorf, Germany). All cross-linking experiments were triplicated for consistency.

2.3. LC-MS/MS analysis

Prior to the LC-MS analysis, digestion mixture was reconstituted with 20 μl of 0.1% formic acid. For analysis, 1 μg of BSA peptide was injected and analyzed with Agilent 1200 series HPLC system equipped with desalting pre-column (Zorbax 300SB-C18 5 μm, 0.3 × 5 mm) followed by analytical column (Zorbax 300SB-C18 3.5 μm, 0.3 × 150 mm from Agilent Technologies (USA)) heated to 60 °C. A flow rate of 10 μl/min with acetonitrile gradient 5%–40% with 35 minutes running time was used. The chromatographic system was directly coupled with electrospray ionization source of SolariX FT-ICR mass spectrometer (Bruker Daltonics, Germany) equipped with 15 T superconducting magnet. Eluted peptides were analyzed in positive broadband mode with 1 M transient data points over the range 250–2500 *m/z*. Data independent mode was used while ESI-TOF tuning mix (Agilent Technologies, USA) served as a lock mass (m/z 922.0098) and as a center of isolation window of ± 500 Da for MS/MS. Fragmentation data were acquired with 22.5 eV collision voltage, 0.2 s MS/MS ion accumulation and 2 accumulated scans per spectra.

2.4. Cross-links identification

In all the spectra, complete chromatograms were deconvoluted with SNAP 2.0 algorithm integrated in Data Analysis 4.4 (Bruker Daltonics, Germany) and exported to the mascot generic files. Obtained files were searched by StavroX 3.6.0.1 [18] with the following settings. Protease sides and modification: C-terminal cleavage of lysine, arginine and tyrosine with a maximum of 3 missed cleavages; static modification carbamidation of cysteines; variable oxidation of methionines. For all experiments, cross-linker was defined as $C_5H_4O_2$ due to the same spacer composition of DSG and BS2G. Cross-linkers specificity was set for N-terminus, lysine, serine and tyrosine. Precursor and fragmentation errors were set to 1.0 and 2.0 ppm, respectively. All cross-linked candidates were manually checked and the data have been deposited to the ProteomeXchange with identifier PXD017299.

2.5. Relative quantification of cross-linked peptides

Ratio of light and heavy linker was determined for all identified cross-linked peptides as previously published [17]. Based on the previously established cross-links from StavroX, the most intense precursor ion (charge state) for cross-linked peptide was selected and used for the creation of extracted ion chromatogram in Data Analysis software. Final spectrum was exported as .XY file by sum of all spectra where precursor ion for light and heavy cross-linked peptide was found. The ratio of light and heavy linker was evaluated in mMass software (version 5.5.0) [19] by fitting experimental data with theoretically simulated isotopic pattern for each cross-linked peptide.

3. Results and discussion

3.1. Chemical cross-linking

The impact of cross-linker polarity on the final outcome was investigated on a well characterized BSA that was widely used as model protein for several cross-linking studies [20]. In this study, BSA was modified by equimolar mixtures of DSG ("light form") with d6-BS2G ("heavy form") and BS2G ("light form") with d6-DSG ("heavy form"). For cross-linking reactions with mixtures composed of DSG/d6-BS2G and d6-DSG/BS2G, 48 and 47 unique cross-links were identified respectively. All of them passed further evaluation, only cross-links confirmed in triplicates were considered and confirmed in mass spectra. Thirty nine cross-links were detected for both chemical probes allowing quantification (Fig. S1). Surprisingly, the observation of isotopic patterns in mass spectra of identified cross-links, revealed that the ratios of light/heavy linked peptides for some cross-links were way far from the expected ratio 50:50. To illustrate the observation of different ratios, two cross-links K211–K242 (Fig. 2A–D) and K221–K294 (Fig. 2E–H) were selected as models. Fig. 2A and E represent spectra where cross-linking mixture of DSG/d6-BS2G was used. However, Fig. 2B and F show reciprocally labeled cross-linking mixture. The use of isotopically labeled cross-linkers enabled monitoring of cross-linking probe origin. In the case of K211–K242 cross-link, the ratio of light/heavy linked peptides revealed preferential formation of cross-linked peptides by non-polar DSG (Fig. 2A, B). On the contrary, cross-link K221–K294 (Fig. 2E, F) showed an opposite trend where modifications mainly came from polar BS2G cross-linker. Additionally, inverted ratio for DSG and BS2G was observed from two experiments where reciprocal mixtures were used.

BSA was also cross-linked with equimolar mixtures of DSG/d4-DSG or BS2G/d4-BS2G to support the previously observed phenomenon. Fig. 2C, D and G, H show "doublet" isotopic patterns for both cross-links with expected ratio of 50:50 for light/heavy form, ensuring the correct sample preparation and chemical activities of the used cross-linking reagents. However, significant differences were observed in absolute intensities of these spectra for non-polar DSG/d4-DSG and polar BS2G/d4-BS2G mixtures. For cross-link K211–K242, the use of non-polar

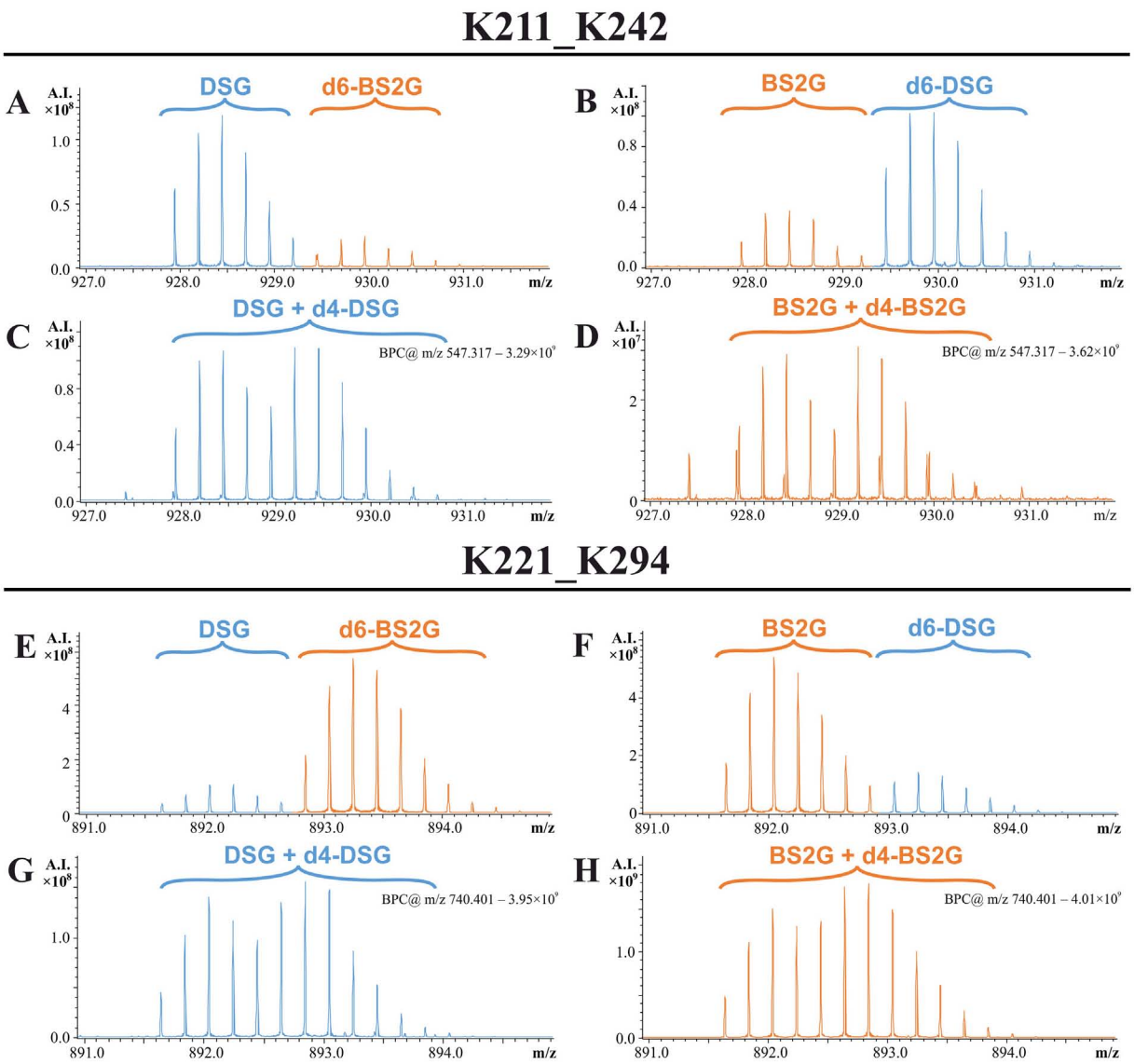


Fig. 2. Detail of mass spectra corresponding to cross-link K211–K242 (A-D) and K221–K294 (E-H) where isotopically labeled equimolar mixtures composed of DSG/d6-BS2G (A, E) and d6-DSG/BS2G (B, F) were used to track the linker origin. For cross-link K211–K242 (A, B) preference of DSG was observed while cross-link K221–K294 (E, F) was predominantly modified by BS2G. Additionally, inverted ratio patterns were observed (A, B) (E, F) for samples were reciprocal mixtures were used. As controls, samples prepared by equimolar mixtures of only one type cross-linker and its four-deuterated analog were prepared. For cross-link K211–K242 (C, D) and K221–K294 (G, H) both control spectra resulted in "doublets" where ratio of light/heavy form was close to ideal 50:50 ratio.

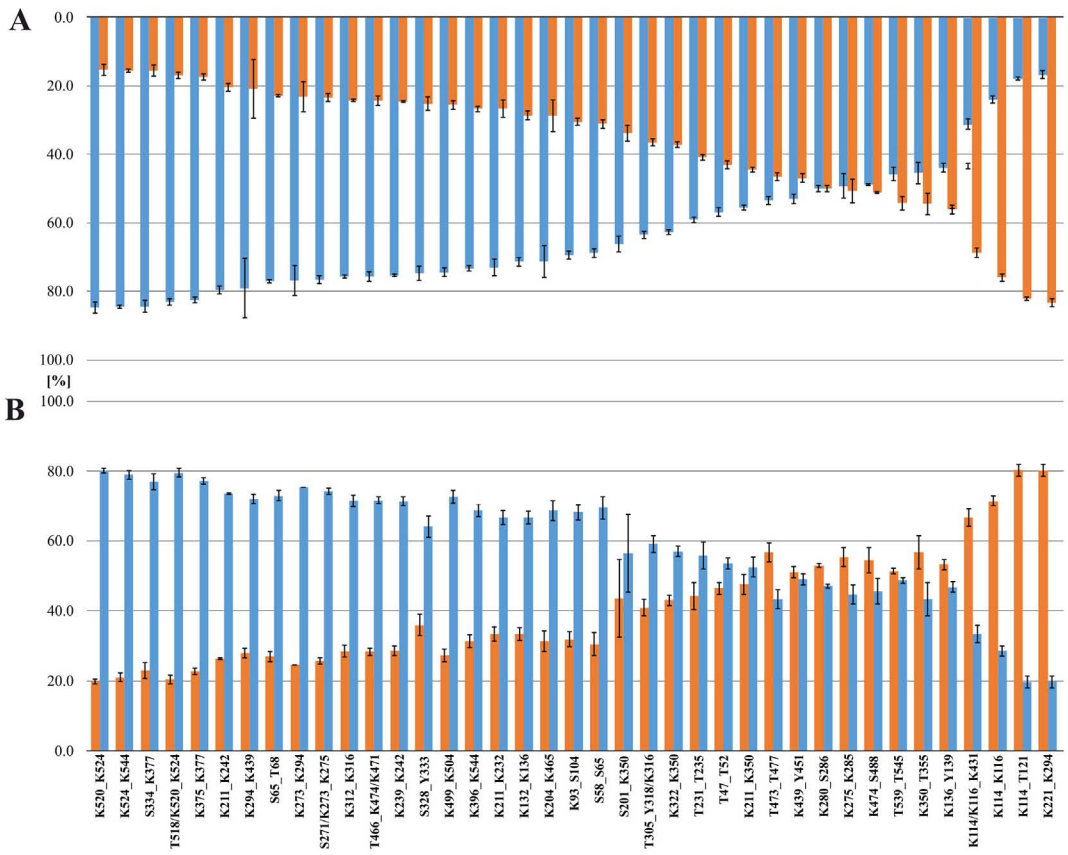


Fig. 3. Relative quantification of non-polar (orange) and polar (blue) cross-linked peptides for two samples prepared with reciprocally labeled mixtures. Fig. A represent sample labeled with DSG/6d-BS2G while fig. B sample cross-linked with mixture of d6-DSG/BS2G. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mixture DSG/d4-DSG (Fig. 2C) resulted in 3.5× higher absolute intensity when compared to the polar mixture BS2G/d4-BS2G (Fig. 2D). This observation is in agreement with the previous results where DSG/d6-BS2G and reciprocal mixtures demonstrated preference for non-polar DSG over K211–K242 cross-link. The same trend was observed for K221–K294 cross-link where the use of BS2G/4d-BS2G (Fig. 2G) resulted in more than 10× higher absolute intensity when compared with non-polar DSG/d4-DSG (Fig. 2H).

3.2. Quantification of cross-links

From the 40 observed cross-links (Table S1) in both samples DSG/d6-BS2G and d6-DSG/BS2G, 39 were quantified for further investigation. The result of quantification is the ratio representing preference for the given cross-link to non-polar DSG or polar BS2G cross-linker. Fig. 3A and B represent the bar plot showing all cross-links with their calculated ratios for DSG/d6-BS2G and d6-DSG/BS2G respectively. Cross-links were sorted by the increase in polarity with less polar (left side) to the most polar (right side). The observed reciprocal pattern between Fig. 3A and B demonstrated the credibility of sample preparation and interpretation.

According to Fig. 3A and B, half of the cross-links preferred non-polar DSG cross-linker, and only 4 cross-links preferred polar BS2G. Overall, only 36% remained close to the 50:50 ratio. Indeed, control experiments revealed 50:50 ratios of light/heavy form over all 39 quantified cross-links when mixtures of DSG/d4-DSG and BS2G/d4-BS2G were used (bar charts available in supporting material as Fig. S2).

3.3. BSA structure

To explain various preferences for both cross-linking probes, it is necessary to consider the protein's chemical properties. From several reported crystal structures of BSA, the structure with the highest resolution (PDB: 4F5S) [21] was chosen for data visualization. To evaluate the polarity regions of protein, color scheme based on amino acid hydrophobicity [22] was applied to the BSA structure where red and white colors represent the hydrophobic amino acids and the hydrophilic residues, respectively. Protein was visualized in surface representation with desired transparency (Fig. 4), and all quantified cross-links were highlighted on structure connecting Cα–Cα atoms of cross-linked residues. Even the preferences of DSG and BS2G probes are clearly visible (Fig. S3), for simplicity, only two cross-links K524–K544 and K116–K431 were chosen to demonstrate the observed phenomena. K524–K544 and K116–K431 were selected as an example for non-polar DSG preferred cross-link and polar BS2G preferred cross-link, respectively. In the case of K524–K544 (Fig. 4A), both cross-linked residues were in close proximity to many hydrophobic amino acids. Due to the restricted accessibility of polar BS2G cross-linker, K524 and K544 were preferentially connected to less polar DSG cross-linker. Unlike the previous example, cross-link K116–K431 (Fig. 4B) was situated in polar region where both cross-linked residues were surrounded by hydrophilic amino acids, thus making this part of the protein more accessible to polar BS2G cross-linker. Therefore, significant preference for BS2G cross-linker over DSG analog was expected and subsequently confirmed by our quantitative cross-linking experiment.

3.4. Venn diagrams

Based on the position in the protein structure, few cross-links were predominantly favored by non-polar or polar cross-linker as shown in Figs. 3 and 4. Hence, the use of single cross-linker (DSG or BS2G) could provide inconsistent results and non-preferential cross-links could be

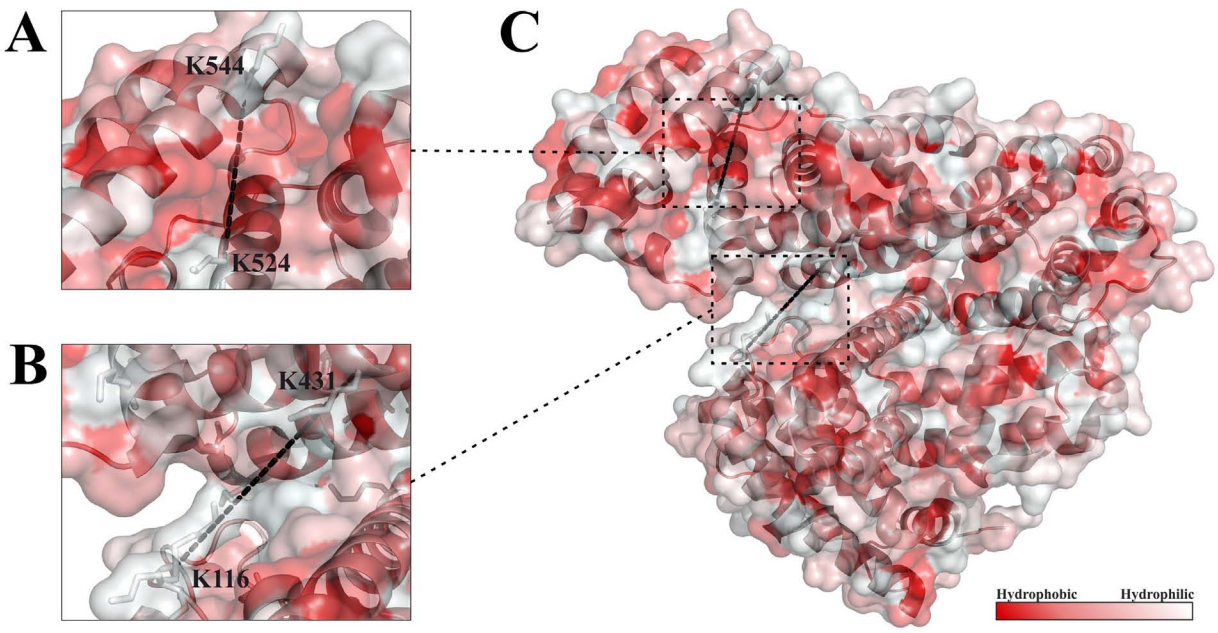


Fig. 4. Model of BSA (PDB entry: 4F5S) with selected polar and nonpolar cross-links. Based on the color scheme cross-link K524–K544 (B) is situated near hydrophobic amino acids (red surface color) thus preferential modification by DSG was observed in this region. On contrary, cross-link K116–K431 is located in polar surface region which makes it more accessible for polar BS2G cross-linker. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

omitted due to their low abundance. Therefore, additional experiments were designed to establish this phenomenon. BSA was cross-linked by non-labeled DSG, BS2G and added with equimolar mixture of DSG/BS2G, keeping reaction conditions, data analysis and interpretation identical in all three cases as well as same as that of previous experimental setup. Fig. 5 represents positively identified cross-links in Venn's diagrams for triplicates using DSG, BS2G and its mixture (list of common cross-links for each triplicate is available in Table S2). In principle, non-polar DSG (Fig. 5A) and polar BS2G (Fig. 5B) enabled the detection of 47 and 46 cross-links, respectively. However, equimolar mixture of DSG/BS2G (Fig. 5C) provided 67 common cross-links (Fig. S4).

In agreement with our quantitative experiments, additional cross-links corresponded to the distance constraints that were not present in samples where only single polarity cross-linking probe was used. Therefore, usage of cross-linker mixture gained the number of identified cross-links for two reasons: some cross-links were selective for polar or non-polar cross-linking probe, and the significant increase in signal intensities facilitated successful cross-link identification in situations where the cross-links were formed by both chemical probes (polar and non-polar).

4. Conclusions

In this article, we have demonstrated that the polarity of cross-linker leaving group had a significant impact on the cross-linking results due to the difference in accessibility of cross-linking agent to reactive residues. This effect was shown on BSA protein where the origin of cross-linker was tracked by the use of isotopically labeled mixtures of DSG/d6-BS2G and d6-DSG/BS2G. Relative quantification of cross-linked peptides revealed cross-links, which were predominantly modified by non-polar DSG or polar BS2G. Additionally, preferences were correlated to the BSA structure, where polar regions were predominantly modified by polar BS2G cross-linker and vice versa. To support our findings, BSA was cross-linked with non-labeled DSG, BS2G and DSG/BS2G equimolar mixture. The DSG/BS2G mixture provided 20 and 21 extra cross-links compared to the experiments where only non-polar DSG and polar BS2G were used, respectively.

The minimal number of distance constraints required for ab initio protein structural model design was one restraint per ten amino acids [4,5]. In many cases, it was hard to reach the desired number using cross-linker targeting primary amino groups, even with the higher amounts of lysine residues. Indeed, the number of available residues, and mainly the lysine accessibility and reactivity, defined the success of cross-linking reaction [23]. Considering these limitations, it was suggested to block the most reactive lysine residues by NHS-acetate prior to the cross-linking reaction or increase the extent of cross-linker over the protein [24]. This idea became obsolete since excessive cross-linking or introduction of negative charge on lysine induced structural changes and potentially led to the creation of artificial cross-links [13]. To avoid this effect, the mixture of non-polar/polar cross-linker in slight excess could be used within structural perturbation.

Taken together, these findings highlighted the role of cross-linker polarity for designing protein cross-linking experiments and provided new insights into sample preparation where cross-linker polarity played an important role.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2020.103716.

Author contributions

P. Novák planned the experiments and supervised the manuscript writing. J. Fiala performed the chemical cross-linking experiments, prepared samples for MS measurements, analyzed acquired data including quantitation and wrote the manuscript. Z. Kukačka helped with the design of chemical cross-linking experiments and participated in MS data interpretation.

Declaration of Competing Interest

The authors declare that they have no conflicts of interests relating to the contents of this article.

Acknowledgment

This work was financially supported by the Czech Science

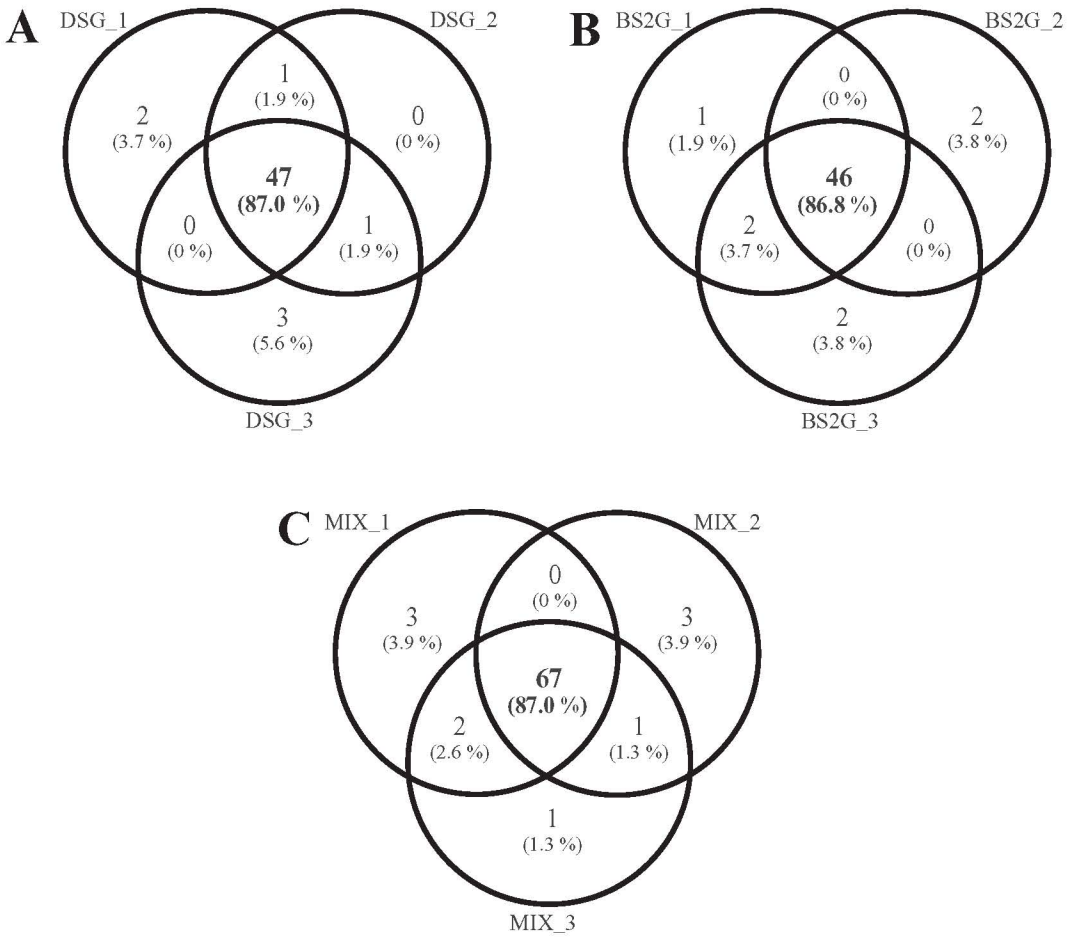


Fig. 5. Venn's diagrams of triplicates for experiments with non-labeled cross-linkers. Mixture of DSG/BS2G (C) resulted in more identified cross-links in one experiment compared to the use of only DSG (A) or BS2G (B) cross-linker per reaction.

Foundation (grant numbers 19-16084S), Charles University (Central Europe Leuven Strategic Alliance - 18/044), the Ministry of Education of the Czech Republic (project LD15089 and program "NPU II" project LQ1604), and European Commission H2020 (European Network of Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry Centers - project agreement No.731077 and European Proteomics Infrastructure Consortium providing Access - project agreement No.823839) and, in part, by the Czech Academy of Sciences (RVO61388971). Also, we acknowledge European Regional Development Funds (CZ.1.05/1.1.00/02.0109 BIOCEV) and the CMS Core Facility at BIOCEV supported by the Czech Infrastructure for Integrative Structural Biology (LM2015043 CIISB for CMS BIOCEV funded by MEYS CR).

References

[1] M.M. Young, N. Tang, J.C. Hempel, C.M. Oshiro, E.W. Taylor, I.D. Kuntz, B.W. Gibson, G. Dollinger, High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 97 (11) (2000) 5802–5806.
[2] T.B. Farmer, R.M. Caprioli, Assessing the multimeric states of proteins: studies using laser desorption mass spectrometry, Biol. Mass Spectrom. 20 (12) (1991) 796–800.
[3] V. Rossi, C. Gaboriaud, M. Lacroix, J. Ulrich, J.C. Fontecilla-Camps, J. Gagnon, G.J. Arlaud, Structure of the catalytic region of human complement protease C1s: study by chemical cross-linking and three-dimensional homology modeling, Biochemistry 34 (22) (1995) 7311–7321.
[4] T.F. Havel, G.M. Crippen, I.D. Kuntz, J.M. Blaney, The combinatorial distance geometry method for the calculation of molecular conformation II. Sample problems and computational statistics, J. Theor. Biol. 104 (3) (1983) 383–400.
[5] T.F. Havel, I.D. Kuntz, G.M. Crippen, The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem, J. Theor. Biol. 104 (3) (1983) 359–381.
[6] J. Rappsilber, The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, J. Struct. Biol. 173 (3) (2011) 530–540.
[7] A. Sinz, The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks, Expert Rev. Proteom. (2014) 733–743 December 16.
[8] G.H. Kruppa, J. Schoeniger, M.M. Young, A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry, Rapid Commun. Mass Spectrom. 17 (2) (2003) 155–162.
[9] P. Novak, G.H. Kruppa, Intra-molecular cross-linking of acidic residues for protein structure studies, Eur. J. Mass Spectrom. (Chichester, Eng) 14 (6) (2008) 355–365.
[10] R.B. Jacobsen, K.L. Sale, M.J. Ayson, P. Novak, J. Hong, P. Lane, N.L. Wood, G.H. Kruppa, M.M. Young, J.S. Schoeniger, Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry, Protein Sci. 15 (6) (2006) 1303–1317.
[11] M. Suchanek, A. Radzikowska, C. Thiele, Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells, Nat. Methods 2 (4) (2005) 261–267.
[12] A.F. Gomes, F.C. Gozzo, Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS, J. Mass Spectrom. 45 (8) (2010) 892–899.
[13] D. Rozbeský, M. Rosůlek, Z. Kukačka, J. Chmelík, P. Man, P. Novák, Impact of chemical cross-linking on protein structure and function, Anal. Chem. 90 (2) (2018) 1104–1113.
[14] D.M. Schulz, C. Ihling, G.M. Clore, A. Sinz, Mapping the topology and determination of a low-resolution three-dimensional structure of the Calmodulin-Melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes, Biochemistry 43 (16) (2004) 4703–4715.
[15] D. Rozbesky, Z. Sovova, J. Marcoux, P. Man, R. Ettrich, C.V. Robinson, P. Novak, Structural model of lymphocyte receptor NKR-P1C revealed by mass spectrometry and molecular Modeling, Anal. Chem. 85 (3) (2013) 1597–1604.
[16] J. Rappsilber, S. Siniossoglou, E.C. Hurt, M. Mann, A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry, Anal. Chem. 72 (2) (2000) 267–275.
[17] Z. Kukacka, M. Rosulek, M. Strohalm, D. Kavan, P. Novak, Mapping protein structural changes by quantitative cross-linking, Methods 89 (2015) 112–120.
[18] M. Götze, J. Pettelkau, S. Schaks, K. Bosse, C.H. Ihling, F. Krauth, R. Fritzsche,

U. Kühn, A. Sinz, StavroX-A software for analyzing crosslinked products in protein interaction studies, J. Am. Soc. Mass Spectrom. 23 (1) (2012) 76–87.
[19] M. Strohalm, D. Kavan, P. Novák, M. Volný, V. Havlíček, MMass 3: a cross-platform software environment for precise analysis of mass spectrometric data, Anal. Chem. 82 (11) (2010) 4648–4651.
[20] C. Iacobucci, C. Piotrowski, R. Aebersold, B.C. Amaral, P. Andrews, K. Bernfur, C. Borchers, N.I. Brodie, J.E. Bruce, Y. Cao, et al., First community-wide, comparative cross-linking mass spectrometry study, Anal. Chem. 91 (11) (2019) 6953–6961.
[21] A. Bujacz, Structures of bovine, equine and leporine serum albumin, Acta Crystallogr. Sect. D Biol. Crystallogr. 68 (10) (2012) 1278–1289.
[22] D. Eisenberg, E. Schwarz, M. Komaromy, R. Wall, Analysis of membrane and surface protein sequences with the hydrophobic moment plot, J. Mol. Biol. 179 (1) (1984) 125–142.
[23] P. Novak, G.H. Kruppa, M.M. Young, J. Schoeniger, A top-down method for the determination of residue-specific solvent accessibility in proteins, J. Mass Spectrom. 39 (3) (2004) 322–328.
[24] X. Guo, P. Bandyopadhyay, B. Schilling, M.M. Young, N. Fujii, T. Aynechi, R.K. Guy, I.D. Kuntz, B.W. Gibson, Partial acetylation of lysine residues improves intraprotein cross-linking, Anal. Chem. 80 (4) (2008) 951–960.

ARTICLE VI

Stalmans G., Lilina A. V., Vermeire P. J., Fiala J., Novák P., Strelkov S. V.

Addressing the Molecular Mechanism of Longitudinal Lamin Assembly Using Chimeric Fusions.

Cells 2020, 9, 1–21.

My contribution: Conducting experiments (chemical cross-linking and LC-MS/MS analysis), data interpretation.

cells

MDPI

Article

Addressing the Molecular Mechanism of Longitudinal Lamin Assembly Using Chimeric Fusions

Giel Stalmans [1,†], Anastasia V. Lilina [1,†], Pieter-Jan Vermeire [1], Jan Fiala [2,3], Petr Novák [2,3] and Sergei V. Strelkov [1,*]

[1] Laboratory for Biocrystallography, KU Leuven, 3000 Leuven, Belgium; giel.stalmans@kuleuven.be (G.S.); anastasia.lilina@kuleuven.be (A.V.L.); pieterjan.vermeire@kuleuven.be (P.-J.V.)

[2] Department of Biochemistry, Charles University, 12800 Prague, Czech Republic; jan.fiala@biomed.cas.cz (J.F.); pnovak@biomed.cas.cz (P.N.)

[3] Institute of Microbiology of the Czech Academy of Sciences, 14220 Prague, Czech Republic

* Correspondence: sergei.strelkov@kuleuven.be; Tel.: +32-1633-0845

† These authors have contributed equally to this work.

Received: 31 May 2020; Accepted: 02 July 2020; Published: 7 July 2020

Abstract: The molecular architecture and assembly mechanism of intermediate filaments have been enigmatic for decades. Among those, lamin filaments are of particular interest due to their universal role in cell nucleus and numerous disease-related mutations. Filament assembly is driven by specific interactions of the elementary dimers, which consist of the central coiled-coil rod domain flanked by non-helical head and tail domains. We aimed to investigate the longitudinal 'head-to-tail' interaction of lamin dimers (the so-called Acn interaction), which is crucial for filament assembly. To this end, we prepared a series of recombinant fragments of human lamin A centred around the N- and C-termini of the rod. The fragments were stabilized by fusions to heterologous capping motifs which provide for a correct formation of parallel, in-register coiled-coil dimers. As a result, we established crystal structures of two N-terminal fragments one of which highlights the propensity of the coiled-coil to open up, and one C-terminal rod fragment. Additional studies highlighted the capacity of such N- and C-terminal fragments to form specific complexes in solution, which were further characterized using chemical cross-linking. These data yielded a molecular model of the Acn complex which features a 6.5 nm overlap of the rod ends.

Keywords: nuclear lamins; intermediate filaments; X-ray crystallography; chemical cross-linking; mass spectrometry

1. Introduction

Lamins represent a distinct class within the intermediate filament (IF) protein family. These nuclear proteins are expressed in all human cell types. Lamin A (LA) and its splice variant lamin C as well as two closely related lamins B1 and B2 jointly form the lamina, a meshwork of ~3.5-nm-thick filaments located at the inner side of the nuclear envelope [1]. By doing this, lamins provide mechanical stability, which is the core function of IF family. Moreover, lamins are involved in a broad variety of cellular processes including chromatin organization and transcription, DNA replication and repair, cell differentiation, mitosis and gene expression [2–7]. As a result, mutations in the lamin genes cause a wide range of diseases called laminopathies, including Emery-Dreifuss muscular dystrophy and Hutchinson-Gilford progeria syndrome [8]. Furthermore, malfunctioning of lamins plays a role in diabetes, heat shock and cancer [9–11].

Like for the whole IF family, the basic constitutive unit of the lamin filament is a rod-like dimer which results from a parallel coiled-coil (CC) structure. In line with this, the primary structure of lamins reveals a central α-helical domain responsible for the CC formation which is flanked by non-helical N- and C-terminal domains called the head and the tail respectively. The central domain is divided into three CC segments known as coil1A, coil1B and coil2, which are interconnected by two linkers, L1 and L12 (Figure 1a). While the length of both coil1A and coil2 segments is almost universally conserved across the IF family, a signature feature of nuclear lamins is the coil1B segment, which is longer by 42 residues (six heptads) compared to cytoplasmic IF proteins [12]. The CC segments are characterized by a pattern of predominantly hydrophobic residues which mostly follow the classical heptad repeat resulting in a left-handed geometry. In addition, coil2 also contains several 11-residue (hendecad) repeats which yield a parallel (untwisted) α-helical bundle [13,14]. Recent crystallographic data revealed that both L1 and L12 linkers in lamins are α-helical even though the CC core is locally interrupted [12,15]. These observations confirmed an earlier suggestion that the linkers represent the points of flexibility within the rod.

In stark contrast with the rod domain, the head and tail domains of IF proteins are dominated by regions predicted to be intrinsically disordered [16]. At the same time, both the head and the tail play an important role in the filament assembly by interacting with specific regions of the rod, and were suggested to accommodate a more ordered structure once the filament is formed [17–19]. In LA, the head domain is relatively short (26 residues) compared to cytoplasmic IF proteins. Specifically in lamins, the tail domain additionally contains a nuclear localization signal and an immunoglobulin-like globular domain (Figure 1a).

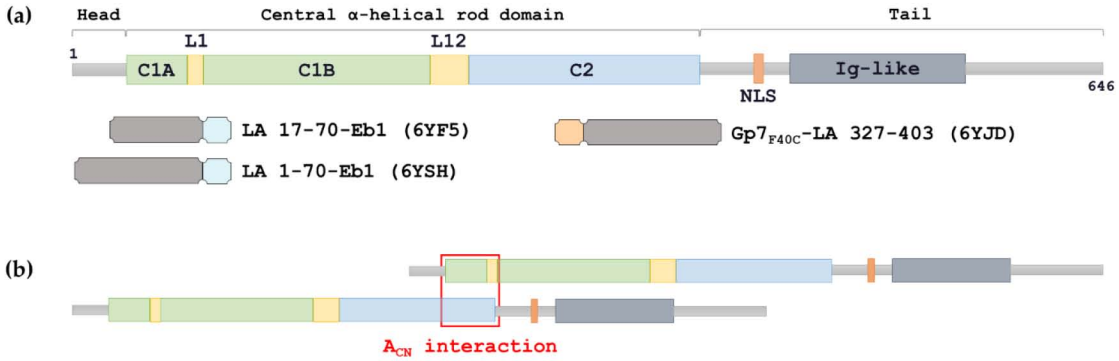


Figure 1. (**a**) Schematic representation of the primary structure of lamin A. C1A, coil1A; L1, linker 1; C1B, coil1B; L12, linker 2; C2, coil2; NLS, nuclear localization signal; Ig-like, immunoglobulin-like domain. Chimeric fragments used here for crystallographic studies are shown below. The capping motifs Eb1 and Gp7$_{F40C}$ are coloured cyan and wheat, respectively. (**b**) Scheme of longitudinal lamin assembly which is based on the A$_{CN}$ interaction.

Assembly of all IF types depends on specific interactions of the elementary dimers. However, the assembly pathway of nuclear lamins is distinct from that of cytoplasmic IFs. Early in vitro studies revealed that lamins can produce longer thin threads of dimers all oriented in the same way, suggesting a longitudinal (so-called A$_{CN}$) interaction of dimers (Figure 1b) as a dominant feature [20–22]. Filamentous structures composed of two antiparallel, laterally associated long dimer threads (thus counting four lamin chains per cross-section) were also present. These results are in line with studies of in vivo assembled lamina using cryo-electron tomography that revealed 3.5-nm-thick filaments [23]. Recent X-ray crystallographic studies have established the molecular detail of the lateral interaction of half-staggered, antiparallel lamin dimers with coil1B segments aligned (so-called A$_{11}$ interaction) [12,15], while additional half-staggered mode (A$_{22}$) aligning coil2 segments was studied using chemical cross-linking [15]. Excessive lateral assembly in vitro, resulting in paracrystals, was also observed [21,24,25]. In general, the appearance of various in vitro assembled

structures was dependent on environmental conditions like pH, ionic strength and presence of calcium [17,20,21,26,27].

In the past, crystallographic studies have been instrumental in developing a better understanding of the IF structure including lamins. While the full-length dimer is too elongated and flexible to be crystallized, shorter rod fragments could be crystallized and resolved [28]. This 'divide-and-conquer' approach has helped to unravel major parts of the rod domain in vimentin, keratins and lamins to atomic resolution. However, when using fragments, it is imperative to provide for the correct formation of parallel, in-register dimeric CCs. Indeed, shorter fragments may not oligomerize at all, form CCs with wrong multiplicity such as trimers, or antiparallel and staggered structures rather than parallel and unstaggered [29]. While also seen for numerous other CC proteins, such complications were documented for fragments of the IF rod domain. For instance, a short fragment of vimentin corresponding to isolated coil1A segment was initially crystallized as a monomer (PDB code 1GK7) and later found to be only marginally stable as a dimer in solution [30,31]. Another example is a C-terminal rod fragment of LA (residues 328–398, PDB code 2XV5) that was found to engage in an unexpected, staggered assembly [32].

To address these problems, stabilization of CC fragments through fusions to other domains was introduced in the past. The idea was that a specific capping motif (N- or C-terminal) would bring together the corresponding ends of the sequence of interest, thereby 'bootstrapping' the formation of the CC. A natural example here is the C-terminal motif, known as the 'foldon', of the trimeric CC protein fibritin from bacteriophage T4 [33]. An early attempt to stabilize the last 28 residues (385–412) of the rod domain of human vimentin involved an N-terminal fusion to the GCN4 leucine zipper [31]. The latter was chosen because it had been known as a prototype CC dimer with high thermal stability [34]. More recently, bacteriophage φ29 scaffolding protein Gp7 [35] and microtubule binding protein Eb1 [36] domains were used as N- and C-terminal caps, respectively. These capping motifs enabled X-ray structure determination of several myosin fragments as well as tropomyosin overlap [37–39].

In the current work, our aspiration was to address the molecular detail of longitudinal lamin assembly (Figure 1b) using X-ray crystallography as the main tool. In the past this question could be tackled by preparing shorter N- and C-terminal fragments of the rod domain (dubbed 'mini-lamins') and examining their interaction in solution [40]. Here we further refined this approach in two directions. First, the cloned N- and C-terminal rod fragments were supplemented with C-terminal and N-terminal capping motifs respectively (Figure 1a), to ensure correct dimer formation. Second, our fragments included adjacent regions of the head and tail domain respectively, since these regions were shown to be important for longitudinal lamin assembly [27,40,41].

As a result, we were able to firstly determine three crystal structures of individual LA fragments. These included two N-terminal fragments comprising residues 1 to 70 and 17 to 70 respectively, both C-terminally fused to the Eb1 domain, and a C-terminal fragment comprising residues 327 to 403, N-terminally fused to the Gp7 domain. The N-terminal lamin fragments in particular reveal a previously unknown β-structural motif located at the very N-terminus of the rod domain, and also provide evidence towards the CC unzipping within the coil1A segment. Second, we were able to study the interaction of N- and C-terminal rod fragments using gel filtration and chemical cross-linking. As a result, we propose a three-dimensional molecular model of the longitudinal A_{CN} interaction of nuclear lamins.

2. Materials and Methods

2.1. Cloning, Expression and Purification

The overall strategy towards cloning and recombinant protein purification was as described in [29]. Initially the DNA sequences for the capping motifs were purchased as gBlocks Gene Fragments from Integrated DNA Technologies (Leuven, Belgium). Prior to cloning, a Quick-Change site-directed mutagenesis was performed to introduce a F40C mutation into the Gp7 cap (see

Supplementary Table 1 for primers). Sequence and ligation-independent cloning was performed using a pETSUK2 vector [42]. The resulting plasmids encoded a 6xHis tag and a small ubiquitin-related modifier (SUMO) domain, followed by the sequence of interest.

Overexpression of obtained constructs (Supplementary Figure 1a) was done in *E. coli* Rosetta 2 (DE3) pLysS strain (Merck, Germany) by auto-induction in the ZYP-5052 medium [43,44]. Cells were harvested by centrifugation, resuspended in low-imidazole buffer (12.5 mM imidazole, 250 mM NaCl, 40 mM Tris HCl pH 7.5, 5 mM βME) containing a lysis mixture (10 mM $MgCl_2$, 1% Triton X-100, 5% SIGMAFAST™ inhibitor cocktail (Sigma-Aldrich, Overijse, Belgium), 1 mM PMSF and 100U Cryonase cold-active nuclease (Takara Bio Europe SAS, Saint-Germain-en-Laye, France)), sonicated and clarified by centrifugation. The supernatant was loaded onto a Ni-chelating column (His60 Ni Superflow resin, Takara Bio Europe SAS, Saint-Germain-en-Laye, France), pre-equilibrated with low-imidazole buffer. 6xHis-SUMO tagged chimeras were trapped on the column and eventually eluted by applying high-imidazole buffer (500 mM imidazole, 250 mM NaCl, 40 mM Tris HCl pH 7.5, 5 mM βME). 6xHis-SUMO tag was cleaved by overnight incubation (4 °C) with SUMO Hydrolase 7K (1:1000 ratio) while dialyzing against a low-imidazole buffer. Afterwards the cleaved mixtures were loaded again onto the Ni column. Chimeric fusions were eluted by low-imidazole buffer while 6xHis-SUMO tags were trapped on the column. Finally, size-exclusion chromatography was performed on a Superdex 200 Increase 10/300 GL column (GE Healthcare Europe, Diegem, Belgium) in 10 mM Tris HCl pH 7.5, 150 mM NaCl. Purified fractions were concentrated using Amicon® Ultra filters with 3kDa cut-off (Merck Millipore, Overijse, Belgium).

2.2. Crystallization and X-Ray Structure Determination

Purified individual proteins were screened for crystallization using commercial kits (Hampton Research, Molecular Dimensions, Qiagen and Rigaku). Subsequently, extensive optimization was performed using a Dragonfly robot (SPT Labtech, Hertfordshire, UK).

$Gp7_{F40C}$-LA 327-403 (9 mg/mL) was crystallized at 4 °C by the hanging drop method using 35% (*v*/*v*) methanol, 0.2 M $MgCl_2$ and 0.1 M HEPES (pH 8.2) as precipitant. Crystals were mounted on cryo-loops using mother liquor supplemented with 30% (*v*/*v*) glycerol. Native data were collected at beamline Proxima-1, Synchrotron Soleil (Saint-Aubin, France). Standard processing in XDS [45] was performed to obtain a complete diffraction data set up to 2.9 Å resolution. Towards phasing by anomalous scattering on sulphur atoms, an additional dataset was collected using the wavelength of 1.8 Å. Three 720° helical scans were performed on different spots of the same large single crystal. The data were merged using XSCALE [45], yielding a redundancy of 120 and a significant anomalous signal up to 3.2 Å resolution.

The anomalous data were submitted to the Auto-Rickshaw pipeline using a web server [46]. Initial search for the positions of anomalous scatterers using SHELXD [47] was followed by processing in Phaser [48], including 12 sites per monomer. Experimentally phased map enabled the initial tracing of the structure, which could be further improved using Buccaneer [49]. Placement of the $Gp7_{F40C}$ cap structure into the initial map helped to determine the right direction of the α-helices forming the CC. Eventually, the strongest anomalous scattering position was attributed to the disulphide bridge formed by the engineered cysteine residue within the $Gp7_{F40C}$ cap. Further five scatterers corresponded to sulphurs present in residues Met278 (Met2 of the $Gp7_{F40C}$ cap), Met345, Met349, Met352, and Met371 (Figure 2a). Of the remaining putative anomalous scatterers, two could be assigned to Ni ions. One of them is coordinated by residue His285. While no metals were present in the crystallization condition, Ni ions could have attached to the protein during purification procedure.

LA 17-70-Eb1 (8 mg/mL) was crystallized at 4 °C by the sitting drop method using 35.7% (*w*/*v*) 1,6-hexanediol, 5% (*w*/*v*) PEG 1K and 0.1 M trisodium citrate dihydrate (pH 4.9). Crystals were mounted on cryo-loops using mother liquor supplemented with 30% (*v*/*v*) glycerol. Native data were collected at beamline I04, Diamond Light Source (Didcot, UK). Data processing to 1.83 Å resolution was performed with xia2 [50] using DIALS [51] for indexing, refinement and integration, POINTLESS

[52] for space group verification and AIMLESS [53] for scaling. Xia2 suggested that usable data extended to 1.83 Å resolution. Initial phasing was performed via molecular replacement using the Eb1 cap (dimer) as search model in Molrep [54], which established the positions of two dimers per asymmetric unit.

LA 1-70-Eb1 (14 mg/mL) was crystallized at 19 °C by the sitting drop method using 0.1 M sodium phosphate monobasic monohydrate, 0.1 M potassium phosphate monobasic, 2.0 M NaCl and 0.1 M MES monohydrate (pH 6.5). Crystals were mounted on cryo-loops without any additional cryo-protectant. Native data to 2.83 Å resolution were collected at beamline P14, Petra III storage ring (Hamburg, Germany), and processed in the same way as for the LA 17-70-Eb1 construct. Phasing could be performed by molecular replacement using a part (starting with residue 35 and containing the full Eb1 cap) of the LA 17-70-Eb1 structure taken as a dimer.

For all structures, Coot [55] was used for interactive model rebuilding. Automated refinement was carried out using Refmac5 [56] and Buster [57]. CC geometry was analysed using TWISTER via an online server (https://pharm.kuleuven.be/apps/biocryst/twister.php) [58]. Pymol (Schrödinger, New York, USA) was used to prepare figures. All statistics on crystallographic structure determination are shown in Supplementary Table 2.

2.3. Size Exclusion Chromatography (SEC) and Multi-Angle Light Scattering (MALS)

N- and C-terminal lamin chimeras were mixed in various molar ratios and incubated for 20 min at 4 °C. Thereafter the individual proteins and the mixtures (50 µL, 1 mg/mL total protein concentration) were run on a Superdex 200 Increase 10/300 GL at 20 °C using an Akta Purifier 10 system (GE Healthcare Europe, Diegem, Belgium). Analysis of the elution fractions was done through sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie staining overnight [59]. Molecular mass determination was performed using an inline SEC-MALS setup. Light scattering was detected using a Dawn Heleos (Wyatt, Santa Barbara, USA) and results were analysed with ASTRA 5 software (Wyatt, Santa Barbara, USA). The experiments were repeated in two different buffers, 10 mM Hepes (pH 7.5) with 150 mM NaCl and 10 mM TEA (pH 7.0) with 150 mM NaCl, which corresponded to those used in different chemical cross-linking experiments (see next section). No difference in SEC elution profiles or MALS-derived molecular masses was observed.

2.4. Chemical Cross-Linking and Mass-Spectrometry (MS) Analysis

Cross-linking experiments were carried out according to Rozbesky et al. [60] with minor modifications. Initial optimization of the cross-linking procedure was through SDS-PAGE. Towards MS-based analysis, the major SEC elution peak corresponding to the complex of LA 17-70-Eb1 and $Gp7_{F40C}$-LA 327-403 (0.4 mg/mL total protein concentration in 10 mM Hepes buffer (pH 7.5) with 150 mM NaCl) was subjected to cross-linking with disuccinimidyl dipropionic urea (DSPU, CF Plus Chemicals, Brno-Řečkovice, Czech Republic) in 1:50 molar excess. SEC-purified complex of LA 22-70-Eb1 and $Gp7_{F40C}$-LA 327-403 (0.4 mg/mL in 10 mM TEA (pH 7.0) with 150 mM NaCl) was cross-linked with disuccinimidyl glutarate (DSG/d6-DSG, molar ratio 1:1, Creative Molecules, Victoria, Canada), disuccinimidyl dibutyric urea (DSBU, CF Plus Chemicals) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, Sigma-Aldrich, Saint Louis, USA) in the molar excess of 1:75, 1:75 and 1:300, respectively. Prior to digestion, disulfide bonds were reduced with 20 mM TCEP at 56 °C for 20 min and alkylated with 20 mM IAA at room temperature for 20 min in the dark. Subsequently, protein samples were diluted five times with 50 mM 4-ethylmorpholine acetate (pH 8.5)/acetonitrile (90:10 *v/v*) and trypsin was added (1:20 ratio). Samples were digested overnight at 37 °C and reaction was quenched by adding trifluoroacetic acid to 0.1%.

The LC-MS/MS analysis was performed as described in [61]. More details are present in Supplementary Text 1. In all the spectra, chromatograms were deconvoluted using SNAP 2.0 algorithm integrated in Data Analysis 4.4 (Bruker Daltonics, Leipzig, Germany) and exported as mascot generic files (mgf). Mgf files were searched by StavroX 3.6.0.1 [62] or Merox 1.6.0.1 [63] with the following settings: cleavage at C-end of Lys, Arg and Tyr with a maximum of 5 missed proteolytic

cleavages, fixed carbamidomethylation of cysteines and variable oxidation of methionines. Cross-linker specificity was set as follows: N-termini, Lys, Ser, Thr and Tyr for DSPU, DSBU and DSG; C- and N-termini, Lys, Glu and Asp for EDC. Error tolerance was set to 1.0 ppm for parent ions and 2.0 ppm for fragment spectra. All cross-linked positions (Supplementary Table 3) were manually checked.

2.5. Molecular Modelling

Prior to modelling, small modifications were made to crystal structures of both N- and C-terminal fusions:

- LA 17-70-Eb1 structure: a symmetric regular CC dimer without kinks was constructed by using chain B starting at residue 27.
- $Gp7_{F40C}$-LA 327-403: the crystallographic model ending with residue 381 was extended by five residues which were ordered in the previously published LA coil2 structure (PDB code 1X8Y [64]).

Initial modelling of the heterotetrameric complex was performed manually in Pymol and Coot by opening up the N-terminal part of LA 17-70-Eb1 dimer and the C-terminal part of the $Gp7_{F40C}$-LA 327-403 dimer and creating a tetrameric overlap. To this end, the structure of the GCN4 leucine zipper core mutant P-LI forming a tetramer (PDB code 1GCL) was used as a template. The length of the overlap was chosen such that:

- The heptad patterns of all four chains were in register just like in the homotetrameric template.
- The model satisfied as many distance restraints corresponding to the experimentally observed cross-links for the complex (Supplementary Table 3) as possible.

Here, residue 27 of LA 17-70-Eb1 was used towards distance restraints for the missing residues in the flexible head region.

After manual docking, the model was refined using the GalaxyRefineComplex tool [65] via the GalaxyWEB server [66] (http://galaxy.seoklab.org/). Details of the modelling process are provided in Supplementary Figure 2.

3. Results

3.1. Design of LA Fragments

Four capped LA fragments were designed, overexpressed in recombinant *E. coli* culture and isolated to >95% purity (Figure 1a, Supplementary Figure 1a). Three N-terminal constructs incorporated residues 1–70, 17–70 and 22–70 of human LA respectively, fused to residues 215–251 of the Eb1 protein (C-terminal cap) [36]. These constructs correspond to the complete or truncated LA head domain followed by the complete coil1A segment (predicted to start with residue 27 and end with residue 67 [14]) and the first three residues of linker L1. In all cases, the fusion was performed in such a way that the heptad repeat pattern was preserved from the lamin part into the capping motif. In addition, a construct comprising residues 1–49 of the Gp7 protein (N-terminal cap) [35] fused by LA residues 327–403 was prepared. Here, the lamin part corresponded to the last 54 residues of coil2 (up to residue 380) followed by a small tail portion. Attachment of the Gp7 cap was done to preserve the hydrophobic pattern of coil2 which includes regular heptads interrupted by a single stutter insert at residues 327–330 (LARE) [64]. The last chimera additionally included a F40C point mutation, located at heptad repeat position 'd' within the Gp7 cap. This mutation was made to introduce a disulphide bridge in order to facilitate experimental crystallographic phasing, since the initial location of a disulphide from anomalous diffraction data is substantially easier than of isolated sulphurs [67]. After protein expression and purification, the disulphide bridge could be confirmed by a non-reducing SDS-PAGE (data not shown).

3.2. Crystal Structure of Gp7$_{F40C}$-LA 327-403

The crystal structure of the Gp7$_{F40C}$-LA 327-403 fusion construct (Figure 2a) was established to 2.9 Å resolution. To this end, experimental phasing on sulphur atoms could be used, which had been facilitated by an engineered disulphide bridge in the dimeric Gp7$_{F40C}$ cap. Crystallographic data reveal good electron density for the capping motif and the lamin part up to residue 381, i.e., exactly the predicted end of the rod domain [14]. The tail domain part (residues 382–403) is disordered in the crystals. Superposition of the Gp7$_{F40C}$-LA 327-403 structure with the previously determined structure of the LA 305-387 fragment (PDB code 1X8Y [64]) (Figure 2b) shows high structural similarity and gives a root mean square deviation (RMSD) of 1.29 Å for 54 Cα-positions (residues 327–380). Interestingly, differences between the two structures are the most noticeable near the N-terminal end of the LA sequence, which corresponds to the stutter (residues 327–330). While the LA 305-387 structure becomes increasingly disordered towards its N-terminus, our fusion with the Gp7$_{F40C}$ cap provides for a proper stabilization of the LA sequence. As a result, the CC geometry of Gp7$_{F40C}$-LA 327–403 (Supplementary Figure 3a) reveals an unwinding that is in line with theoretical expectations for a stutter [58,68].

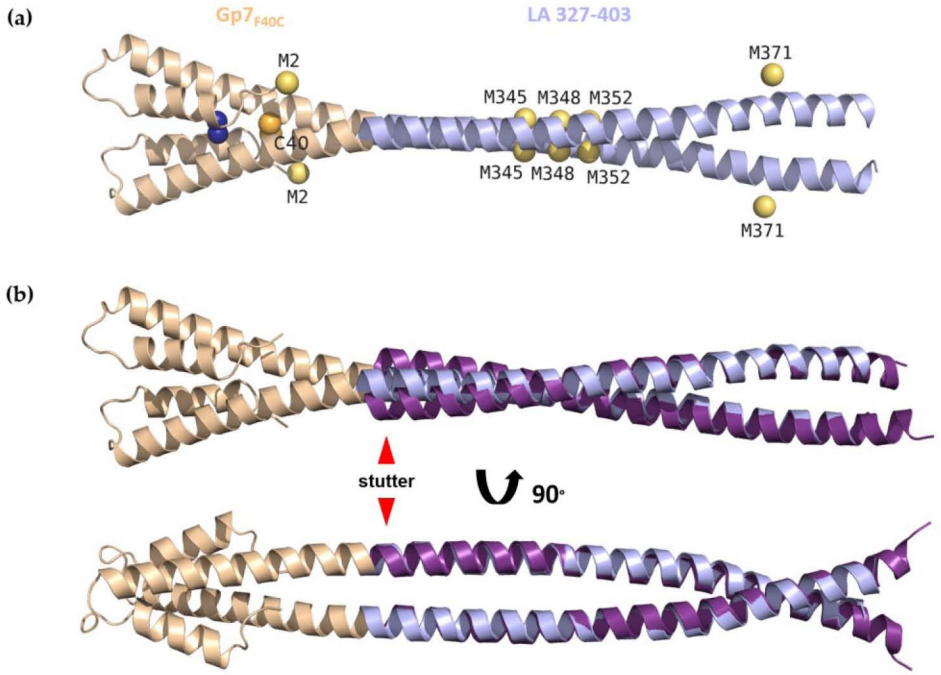


Figure 2. (**a**) Crystal structure of Gp7$_{F40C}$-LA 327-403. The LA 327-403 region and Gp7$_{F40C}$ cap are coloured light blue and wheat respectively. Sulphur atoms used for structure determination are coloured orange (cysteine) and yellow (methionine). Additional bound heavy atom ion (Ni) is coloured deep blue. (**b**) Superposition of the Gp7$_{F40C}$-LA 327-403 structure with residues 327–386 of LA 305-387 structure (PDB code 1X8Y, purple) in two perpendicular views. The stutter (residues 327–330) is indicated by red arrows.

3.3. Crystal Structure of LA 17-70-Eb1 Fragment

Crystallographic data for the LA 17-70-Eb1 fusion construct were phased by molecular replacement, relying on the dimeric Eb1 cap as a search model. The crystal structure was refined to 1.83 Å resolution. The asymmetric unit of the crystal contains two copies of the dimer that are readily superimposable (Figure 3a, Supplementary Figure 4a). In line with the design, a continuous CC with regular left-handed geometry is formed all along the length of the chimera. Importantly, the structure gives a glimpse of the interface between the head domain and coil1A, as reliable electron density starting with residue 23 is seen in all four chains. In addition, chain A is ordered, already beginning

with residue 18, which is due to a crystal contact with a symmetry-related molecule mediated by residue Leu21. The first α-helical residue is Arg28, which is in line with earlier predictions [14,15]. Interestingly, the side chain of Thr27 stabilizes the N-terminus of the α-helix by making a hydrogen bond with the main-chain nitrogen of Gln30 (Figure 3b). Such helical capping function of threonine and serine residues has been described before [69].

The N-terminal region of the dimer reveals two unexpected features. First, residues 23 through 27 corresponding to the proximal residues of the head domain form a β-strand, facilitated by a Pro residue in position 23. Two such strands form a short antiparallel β-sheet, further referred to as the 'β-lock', which flanks the N-terminus of the CC (Figure 3b). The residues Ile26 in the middle of the β-lock are symmetrically interlinked by two main-chain hydrogen bonds, while their hydrophobic side chains are pointing towards the CC.

The second unexpected feature is that, consistently for both dimers in the asymmetric unit, one of the two chains reveals a sharp ~50° change in the α-helical axis direction ('kink') at residue Leu35. Main-chain dihedral angles for this residue are $\varphi = -95°$, $\psi = 0$. While still in the allowed region of the Ramachandran plot, these values are distinct from the typical values ($\varphi = -65°$, $\psi = -45°$) seen for the rest of the α-helix. This kink results in a loss of two standard i->i+4 main-chain H-bonds (between residues Asp34 and Leu38 and between Glu33 and Glu37, respectively). However, either bond is replaced by a H-bond 'bridge' accomplished by an ordered water molecule (Figure 3c). In addition, the standard H-bond between the main-chain oxygen of Lys32 and the nitrogen of Gln36 is in a suboptimal angular position.

Interestingly, the formation of the β-lock and the kink appear to be interdependent. Indeed, the latter causes a drastic deviation of the N-terminus of coil1A from the otherwise typical left-handed supercoiling (Figure 3a). In the absence of the kink the two strands made by residues 23–27 would have been located away from each other (Supplementary Figure 4b).

Coil1A of LA has a pronounced heptad pattern with predominantly hydrophobic residues in 'a' and 'd' positions for the most of its length. However, both the first residue of coil1A (Arg28) corresponding to an 'a' position and the following 'd' residue (Glu31) are not hydrophobic. Instead, polar interactions involving residues Arg28 and Glu31 of one chain and residue Asp34 of another chain (as well as symmetric interactions on the other side of the CC) are observed. The first residue actually creating the hydrophobic core of the dimer is Leu35 (Figure 3a). Our structure superimposes well with the recently published structure of a longer lamin fragment 1-300 (PDB code 6JLB) [15], resulting in a Cα RMSD of 0.94 Å for residues 28–70 (Supplementary Figure 4c). This latter structure consistently features Arg28 as the first α-helical residue and Leu35 as the first residue forming the core of coil1A, while not resolving any residues of the head domain.

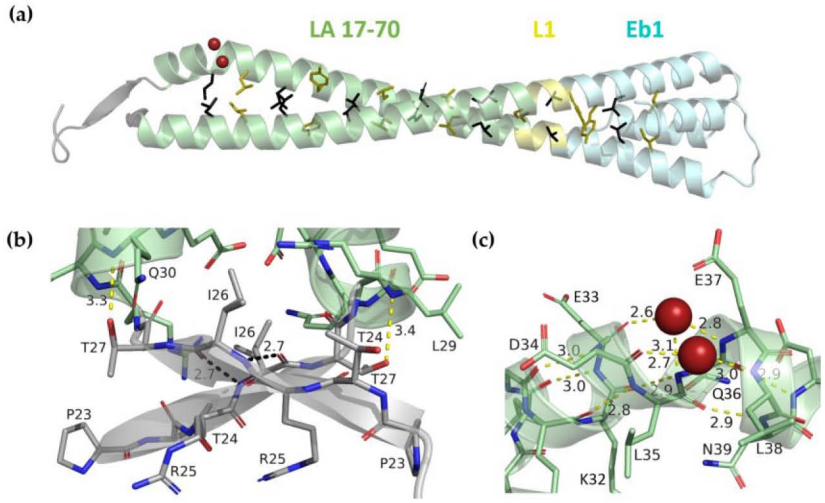


Figure 3. (**a**) Crystal structure of LA 17-70-Eb1. The head residues, coil1A and the first three residues of the linker L1 are coloured grey, pale green and yellow respectively. The Eb1 cap is coloured pale

cyan. Two water molecules involved in H-bond bridges at the α-helical kink are shown as red spheres. Side chains of residues in 'a' (coloured black) and 'd' (coloured olive) positions starting with Leu35 i.e., the residues forming the hydrophobic core of the CC are shown as sticks. (**b**) Top view of the N-terminal end of the dimer featuring an antiparallel β-lock stabilized by two H-bonds (cyan dashed lines) between Ile26 residues. The H-bond between the side-chain oxygen of Thr27 and the main-chain nitrogen of Gln30 is shown (yellow dashed line). (**c**) Zoom in on the α-helical kink present in chain A. Regular main-chain hydrogen bonding pattern as well as two water-based H-bonding bridges are shown (yellow dashed lines). The suboptimal angular position of the H-bond between the main-chain oxygen of residue Lys32 and the nitrogen of residue Gln36 is shown as a purple dashed line.

3.4. Crystal Structure of LA 1-70- Eb1

In addition, we determined the crystal structure of the LA 1-70-Eb1 construct at 2.83 Å resolution. One dimer per asymmetric unit of the crystals is observed. Here the head domain is almost entirely disordered, as reliable electron density only starts from residue 25 (Figure 4). The structure shows a remarkable opening-up of the N-terminal part of the dimer extending up to residue 45. Upon application of the crystal symmetry, two 'unzipped' coil1A dimers form a 3-nm-long antiparallel N-terminal overlap (Figure 4). The overlap features a common hydrophobic core formed by residues Ile26, Leu38, Leu42, Tyr45, Ile46 and Val49 and a symmetric pair of salt bridges (Glu37|Arg41 and Arg41|Glu37) at each side of the overlap. The conformation and interactions of the N-terminal part of the dimer in the context of fusions LA 1-70-Eb1 and LA 17-70-Eb1 are thus entirely different. While in LA 17-70-Eb1 the entire coil1A forms a regular CC structure, which is flanked by the β-lock at the N-terminal end, in LA 1-70-Eb1 all these N-terminal interactions are lost and replaced by the antiparallel contact observed.

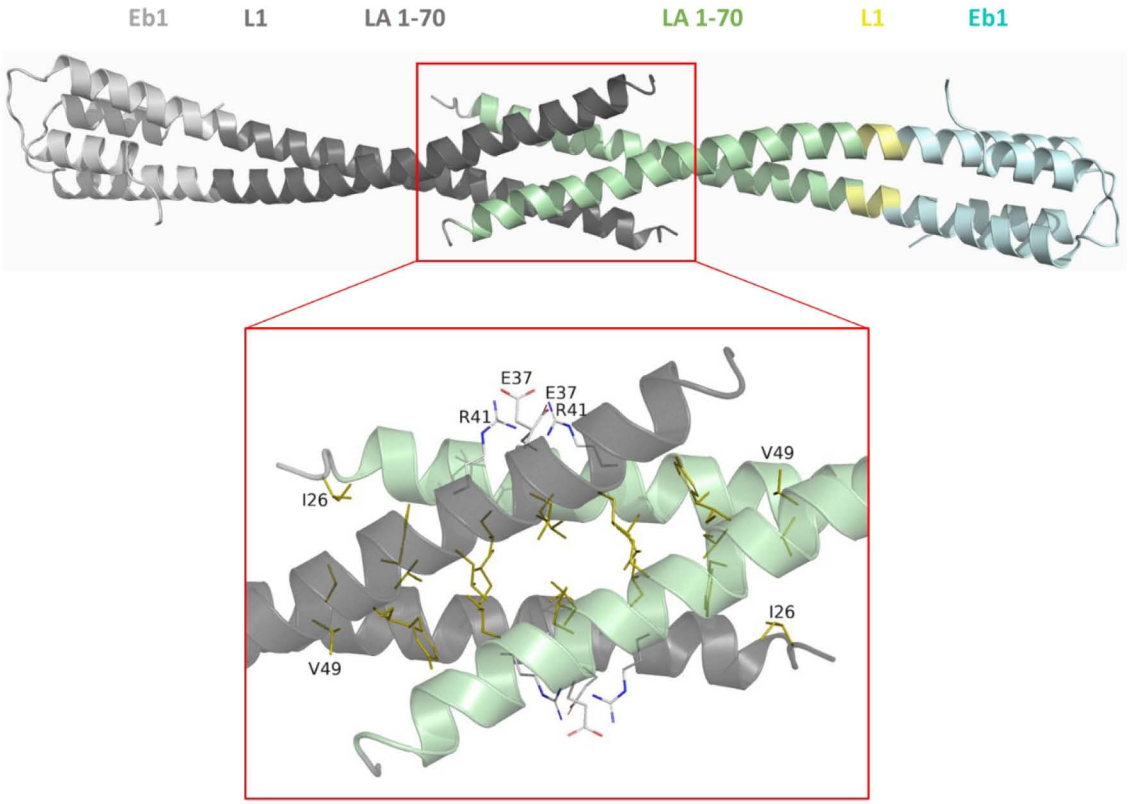


Figure 4. Crystal structure of LA 1-70-Eb1 dimer with coil1A in green, L1 region in yellow and the Eb1 cap in cyan. Additionally, a crystal symmetry mate is shown in black (lamin part) and grey (Eb1 cap). The zoomed area illustrates the interactions responsible for the antiparallel overlap of the N-

terminal parts. Side chains forming the common hydrophobic core (coloured gold), as well as residues Glu37 and Arg41 forming salt bridges (coloured light grey) are shown.

The large differences observed in the dimer structure and crystal lattice contacts between the LA 17-70-Eb1 and LA 1-70-Eb1 structures are most likely due to the different crystallization conditions, rather than a consequence of the complete head being retained in the latter case. Indeed, since the head domain in the LA 1-70-Eb1 structure is almost entirely disordered, it does not seem likely that it plays a major role in the formation of the antiparallel overlap of two coil1A dimers. At the same time, the observed structure clearly reveals the weakness and plasticity of the N-terminal part of coil1A. Indeed, unzipping of the CC has also been observed near the N-terminus of isolated coil1A fragment (residues 102–138) of human vimentin, even in the presence of a stabilizing mutation Y117L [30]. Sequence conservation between the coil1A regions of human LA and vimentin is rather high (70% similarity, see also Supplementary Figure 1c). In fact, the N-terminal opening-up of coil1A dimer is the most pronounced in the LA 1-70-Eb1 structure, followed by a smaller opening in the vimentin structure and a completely regular CC geometry in the 17-70-Eb1 construct (Figure 5).

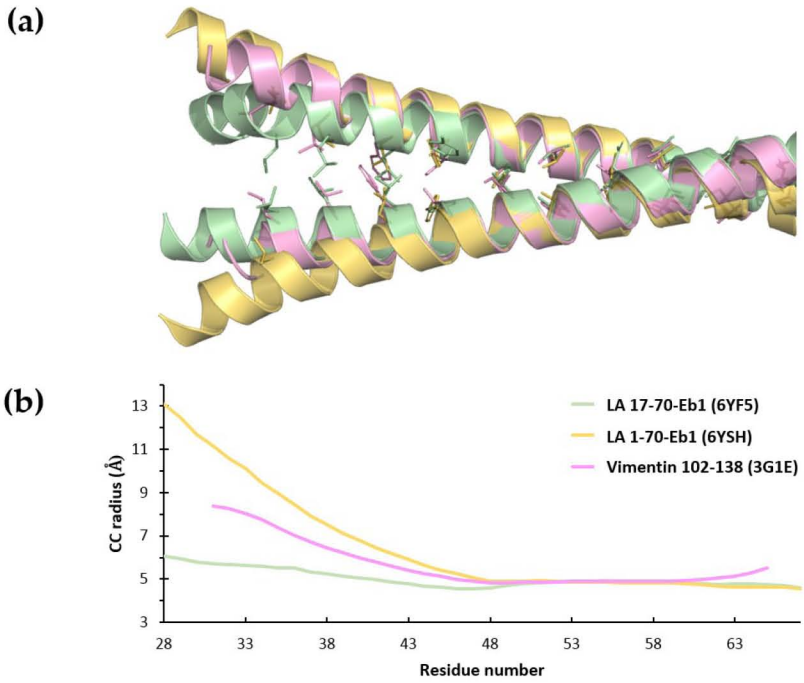


Figure 5. Analysis of the coil1A dimer opening in various structures (PDB codes are shown in brackets). (**a**) Superposition of the coil1A segments seen in LA 17-70-Eb1 (green), LA 1-70-Eb1 (gold) and vimentin 102-138 fragment with mutation Y117L (pink; PDB code 3G1E [30]). The residues in 'a' and 'd' positions forming the hydrophobic core are shown as sticks. (**b**) CC radius plotted as a function of residue number (according to the LA sequence) for the coil1A regions in the three structures, as determined using TWISTER [58].

3.5. Interaction of the N- and C-Terminal Chimeric Constructs in Solution

By design, our N- and C-terminal chimeras are very suited for examining the A_{CN} interaction of lamin dimers. As the first step, we studied their association using SEC. After $Gp7_{F40C}$-LA 327-403 had been mixed in equimolar ratio with either LA 1-70-Eb1, LA 17-70-Eb1 or LA 22-70-Eb1, a single major chromatographic peak was revealed, eluting much earlier than any individual component (Figure 6a). Importantly, SDS-PAGE analysis of multiple chromatographic fractions across the peak revealed both individual proteins in a constant 1:1 ratio (Figure 6b). This pointed to the formation of a higher-order complex between the two constructs, presumably corresponding to the A_{CN} tetramer.

In addition, inline MALS analysis of the SEC elution peaks was performed. Molecular weights (MW) measured for all individual fragments i.e., LA 17-70-Eb1 (21.3 kDa), LA 22-70-Eb1 (19.6 kDa) and Gp7$_{F40C}$-LA 327-403 (28.6 kDa) were in excellent agreement with the theoretical values for the dimers (21.2 kDa, 20.2 kDa, 29.7 kDa, respectively; Figure 6a). However, MW (value at the top of the peak) of the complexes was considerably lower than expected for a heterotetramer with 2:2 stoichiometry. For the LA 17-70-Eb1 / Gp7$_{F40C}$-LA 327-403 complex, the measured MW was 35.2 kDa, compared to a theoretical value of 50.9 kDa. For the LA 22-70-Eb1 / Gp7$_{F40C}$-LA 327-403 complex, the experimental MW was 32.4 kDa, compared to a theoretical value of 50 kDa. It should be noted that the elution peak of the complex was quite broad, while the measured MW value decreased considerably across the peak (Figure 6a). Another important observation is the persistent presence of a minor peak corresponding to the free N-terminal construct in the elution profile of the complex. This minor peak was seen in the SEC profiles of both LA 22-70-Eb1 / Gp7$_{F40C}$-LA 327-403 (after injecting an equimolar mixture of both constructs) and LA 17-70-Eb1 / Gp7$_{F40C}$-LA 327-403 (after injecting the major peak fraction from a previous SEC run; Figure 6a).

These observations suggest that the interaction between the N- and C-terminal lamin constructs is relatively weak, meaning that there is a dynamic equilibrium between the complex and individual components, resulting in a broad elution peak. Of note, the complex eluted at a distinctly earlier position than each of the isolated dimers, despite the apparent average MW for the complex being only ~20% higher than that of the Gp7$_{F40C}$-LA 327-403 dimer. Such elution indicates that the Stokes radius (and therefore at least one dimension) of the complex is considerably larger than that of each individual dimer.

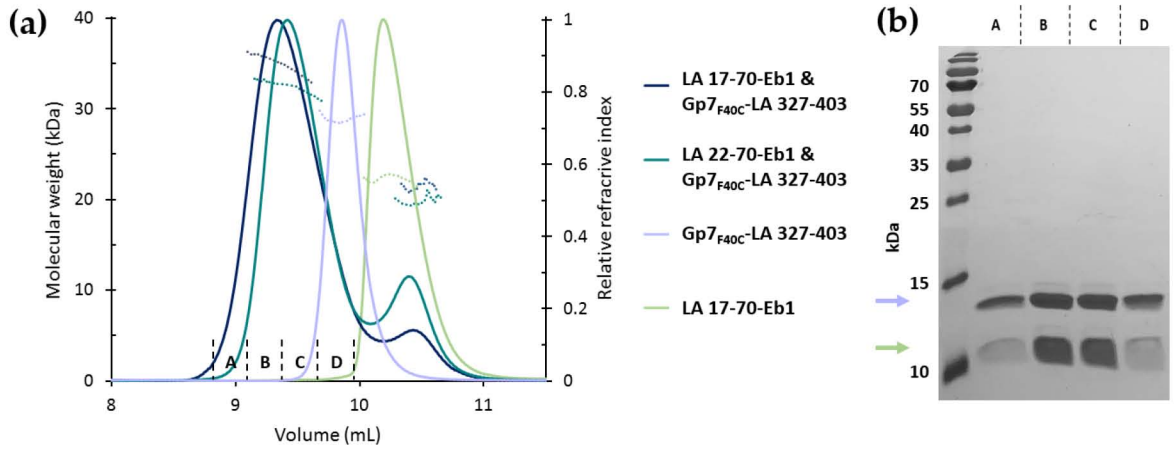


Figure 6. (**a**) SEC-MALS analysis of the LA 17-70-Eb1 / Gp7$_{F40C}$-LA 327-403 complex (1:1 ratio), LA 22-70-Eb1 and Gp7$_{F40C}$-LA 327-403 complex (1:1 ratio), the Gp7$_{F40C}$-LA 327-403 construct and the LA 17-70-Eb1 construct. The LA 17-70-Eb1 / Gp7$_{F40C}$-LA 327-403 complex and both individual constructs were run in 150 mM NaCl, 10 mM Hepes (pH 7.5) buffer. The second complex was run in 150 mM NaCl, 10 mM TEA (pH 7) buffer. Solid lines show the normalized refractive index profiles. Dashed lines show the MALS-based molecular weight values across the peaks. (**b**) SDS-PAGE of chromatographic fractions corresponding to the main peak of the LA 17-70-Eb1 / Gp7$_{F40C}$-LA 327-403 complex in (a). The blue and green arrows indicate Gp7$_{F40C}$-LA 327-403 and LA 17-70-Eb1, respectively.

The SEC-purified complexes of N- and C-terminal constructs were extensively screened for crystallization. Several crystal forms could be obtained. However, all crystals were found to be produced by the Gp7$_{F40C}$-LA 327-403 fusion alone, as evident from SDS-PAGE analysis of the crystals. Moreover, preliminary X-ray characterization always yielded the same space group and cell parameters as seen for Gp7$_{F40C}$-LA 327-403 alone. We hypothesize that the presence of crystallization agents was disrupting the complex, which we have shown to be relatively weak.

3.6. Chemical Cross-Linking

As the next step, both individual N- and C-terminal fusions as well as SEC-purified complexes thereof were subjected to chemical cross-linking using three monofunctional cross-linkers with primary reactivity against amino groups (DSPU, DSBU and DSG) as well as a heterobifunctional cross-linker EDC with reactivity against amino and carboxy groups. After optimization of the procedure, SDS-PAGE analysis revealed a range of cross-linked products in all samples. Specifically, for all cross-linkers, the bands corresponding to homodimers (at ~20 and ~30 kDa, respectively) were dominant for both the individual fusions and the complex (Figure 7). In addition, cross-linking the complex using EDC yielded a distinct band compatible with the 'heterodimer' (i.e., the result of cross-linking of one N-terminal and one C-terminal chain). This band was not apparent for the DSG or DSBU cross-linkers. Importantly, cross-linking of the complex with DSG or DSBU, but not EDC, resulted in a distinct band with a mass close to 50 kDa, which corresponds to a heterotetramer composed of two N-terminal and two C-terminal fragments, as well as two lower bands that could be interpreted as two types of heterotrimers. No higher bands than the heterotetramer were observed (Figure 7), suggesting that the latter is the largest species present in the sample in a prominent concentration.

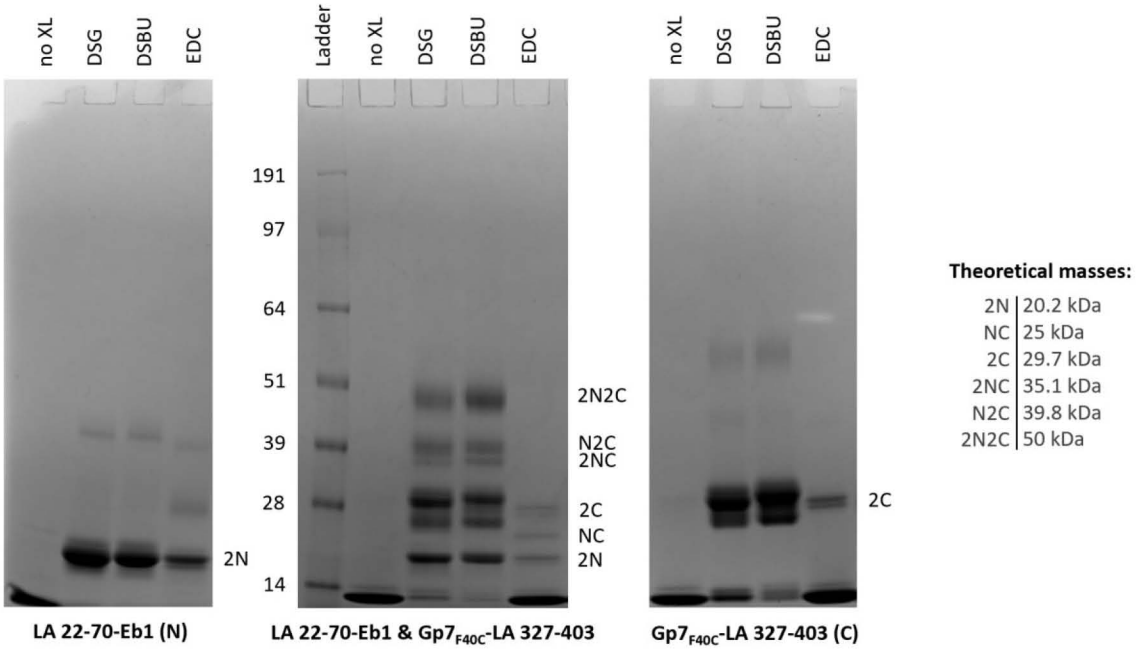


Figure 7. Reducing SDS-PAGE analysis of LA 22-70-Eb1, Gp7$_{F40C}$-LA 327-403 and the SEC-purified complex of these two fragments without cross-linking (no XL) and after cross-linking with DSG, DSBU and EDC. The most likely stoichiometry of different bands is given at the right-hand side of each gel, where letters N and C indicate the N- and C-terminal fragments respectively. Theoretical masses of all bands are given in the side panel.

Next, the complexes of N- and C-terminal fragments were cross-linked and subjected to LC-MS/MS analysis. As the result of experiments involving four different cross-linkers (DSG, DSBU, DSPU and EDC) a total of 51 cross-links were identified (Supplementary Table 3, Supplementary Figure 5). Of those, seven involved the tail portion (not resolved in the crystal structure) of the Gp7$_{F40C}$-LA 327-403 fusion. The remaining 44 cross-links were used in subsequent structural modelling.

3.7. Molecular Modelling of the Acn Complex

Results of chemical cross-linking were supportive of a heterotetramer composed of two N-terminal fragments and two C-terminal fragments, i.e., the species representing the Acn interaction between LA dimers. Overall, this complex should be composed of both the N- and C-terminal dimers in a parallel orientation and with some overlap length, for which several different values were reported in the literature (see Discussion). As a starting hypothesis towards the detailed molecular architecture of the complex, we took a heterotetrameric CC structure, an architecture that has been proposed previously [40,64]. To this end, the crystal structures of the LA 17-70-Eb1 and Gp7$_{F40C}$-LA 327-403 dimers were manually docked together upon opening up the interacting ends of each dimer. The overlapping part was modelled using a parallel homotetrameric heptad-based CC structure as a template. The length of the overlap was varied to satisfy the maximal number of observed chemical cross-links (Supplementary Table 3, Figure 8a,b).

Our optimized Acn tetramer model (Figure 8) has an overlap of 6.5 nm, measured as the distance between the ends of LA rod (i.e., residue 27 within the N-terminal fragment and residue 381 within the C-terminal fragment), projected on the long axis of the heterotetramer. The overlap is thus longer than the length of coil1A alone (4.7 nm). The heterotetramer has a common hydrophobic core involving residues in 'a' and 'd' positions of both N- and C-terminal fragments (Figure 8c).

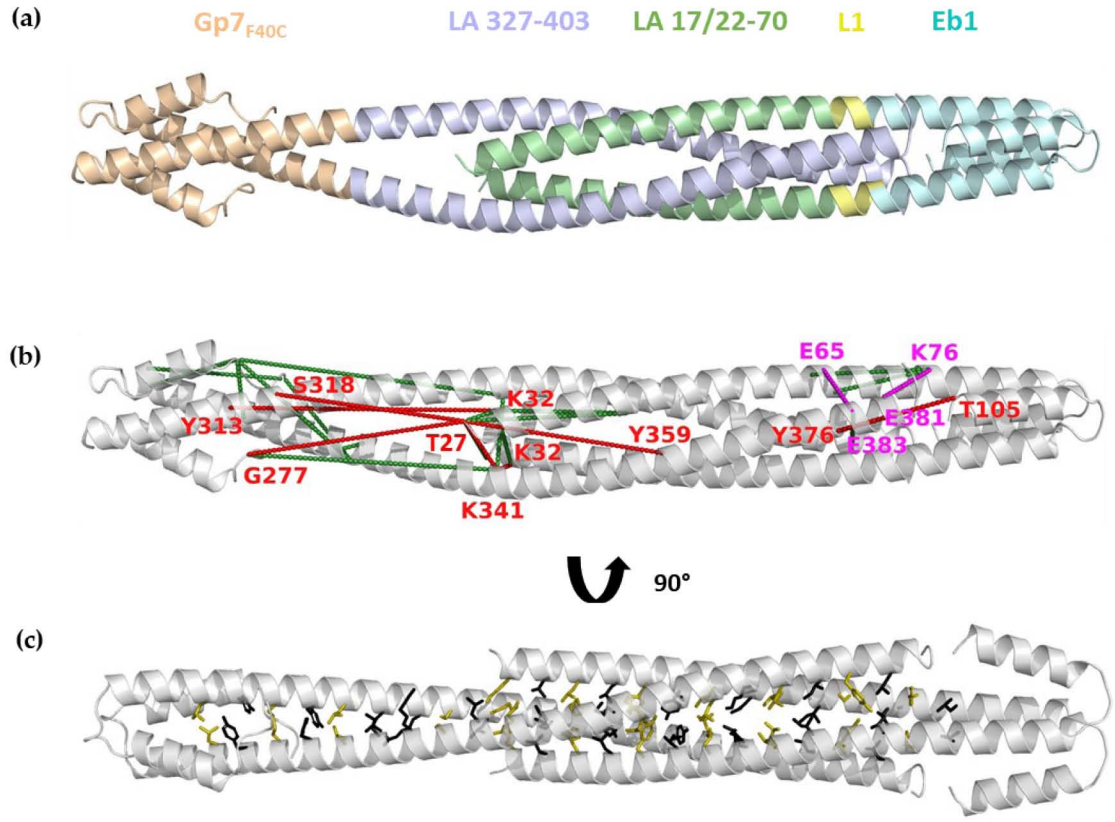


Figure 8. Modelling of the Acn tetramer. (**a**) Final molecular model of the Acn tetramer. Gp7$_{F40C}$-LA 327-403 is coloured wheat (Gp7$_{F40C}$) and blue (LA 327-403). LA 17-70-Eb1 is coloured green/yellow (LA 17-70) and cyan (Eb1). (**b**) Final molecular model of the Acn tetramer with all 31 cross-links that were compatible with the model. Interdimer cross-links are coloured green. Tetramer cross-links are labelled and coloured red (DSBU, DSPU and DSG) and magenta (EDC). (**c**) Rotated view of the model with the hydrophobic core residues shown as sticks ('a' and 'd' positions in black and olive, respectively).

Out of a total of 44 cross-links used, 31 are compatible with the current Acn heterotetramer model, given the maximal allowed distances between the respective Cα positions (Supplementary

Table 3). These include, first, 21 cross-links with both ends corresponding to the same sequence (i.e., either intrachain or intradimer cross-links) and, second, 10 interdimer cross-links (Figure 8b). The latter group includes two EDC-based cross-links between residues Lys76 and Glu381/Glu383 and Glu65 and Lys378, respectively. These zero-length cross-links are particularly useful towards modelling as they provide the most stringent restraint, i.e., a maximum of 15 Å between Cα positions. In our current model, both cross-links correspond to a Cα-Cα distance of 12 Å. Of note, the Glu65-Lys378 cross-link was recently reported for the full-length LA [70].

At the same time, the remaining 13 cross-links (30%) could not be accounted for by the constructed A_{CN} model. Indeed, chemical cross-linking is known to produce artefacts, including linkages resulting from random collisions [71]. In our case, these 13 cross-links could be readily explained by an antiparallel association of two heterotetramers (Supplementary Figure 6). While both our SEC experiments (Figure 6) and SDS-PAGE analysis of cross-linking products (Figure 7) did not explicitly reveal such octameric species, their presence in small concentration cannot be excluded.

The final A_{CN} model was additionally analysed using the program PISA (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html, [72]), which suggested that the obtained assembly is stable in solution (Supplementary Table 4).

4. Discussion

Here we showed that carefully chosen short terminal fragments of lamin rod domain (following the concept of 'mini-lamins' originally proposed in Stuurman et al. [24]) can be conveniently used for studying the longitudinal assembly in molecular detail. A key element of our strategy was the use of relatively short (~50 residues) N- and C-terminal rod fragments that are additionally stabilized by capping motifs (Figure 1). Specifically, the Eb1 capping motif was attached at the C-end of several LA fragments that included a part of the head domain followed by the coil1A segment, and the Gp7 $_{F40C}$ capping motif was attached at the N-terminal end of a LA fragment corresponding to the last few heptads of coil2 and the beginning of the tail domain (Supplementary Figure 1). Such fusions expose the opposite ends of the rod domain in the same way as they would be present in the context of a full-length lamin dimer. Correspondingly, the association of these two fusions in solution should reproduce the A_{CN} interaction between the full-length lamin dimers which is responsible for longitudinal assembly.

As demonstrated here, such fusions provide several benefits. First, the capping motifs used (and especially the Gp7$_{F40C}$ cap that had been additionally stabilized by a disulphide bridge) provide for the correct assembly of dimeric, registered CCs even for short lamin fragments. Second, the capping motifs help to phase the crystallographic data. Indeed, both the Eb1 and the Gp7$_{F40C}$ caps employed here consist of an α-helical region yielding a stable dimeric CC and an additional α-helix folding back onto it. In contrast to a bare CC segment, such cap structure is more suitable as a molecular replacement model, since it does not suffer from 'internal' symmetry (i.e., it does not overlap with itself upon a seven-residue register shift [29]). Following an alternative possibility, an engineered mutation to cysteine in the Gp7$_{F40C}$ cap has facilitated experimental phasing using the anomalous scattering of sulphurs. Finally, in all three solved crystal structures, the capping motifs were involved in lattice contacts (as exemplified in Supplementary Figure 3b,c for Gp7$_{F40C}$-LA 327-403 and in Supplementary Figure 4d,e for LA 17-70-Eb1). These observations suggest a likely beneficial role of including the capping motifs towards crystallization.

The three solved crystal structures contribute to a better understanding of the molecular organization of lamin dimer. In particular, the LA 17-70-Eb1 structure (Figure 3) gives a glimpse of the proximal half of the lamin head domain. Indeed, for two out of four chains in the asymmetric unit, the head part was found to be ordered. Although this was only a result of interactions with symmetry-related dimers in the crystal lattice, this observation supports an earlier hypothesis that the head domains of IF dimers, while being dynamic in isolation, become more ordered upon binding to other dimers [73]. This interaction appears to catalyse the filament assembly. Moreover, our crystal data indicate that the last five residues of lamin head domain can form a β-strand (residues 23–27).

Two such strands yield a 'β-lock' flanking the N-terminus of coil1A, although this requires a formation of an α-helical kink at residue 35 in one of the chains of the dimer. Finally, the N-terminus of the α-helix was stabilized by a hydrogen bond provided by the Thr27 residue, in line with a frequently observed α-helical capping mechanism [69].

Of note, in lamin A, B1 and B2, the primary structure near the border of the head and coil1A domains is highly conserved (Supplementary Figure 1b), suggesting that the β-lock could be a common feature in all lamins. However, lack of sequence conservation in other IF chain types (Supplementary Figure 1c) speaks against a similar β-structure being formed there. For vimentin but not for other IF types, an additional short α-helix (denoted as the 'pre-coil' domain) is predicted upstream of coil1A. Interestingly, both the pre-coil domain and coil1A in vimentin seem to be stabilized by N-terminal capping Thr/Ser residues, just like we observed in LA coil1A.

Importantly, of the two N-terminal LA fragments, the 17-70-Eb1 structure represents a CC with regular geometry, stabilized by the 'β-lock'. In contrast, the 1-70-Eb1 structure features an opened-up, 'unzipped' coil1A segment, even though the presence of the Eb1 cap keeps the two chains together at the C-terminus. The observed unzipping (Figure 4) is especially interesting with respect to the possible mechanism of longitudinal IF assembly. Recently, mutation Y45C in LA which is linked to Emery–Dreifuss muscular dystrophy has been shown to be detrimental to the Acn interaction [15]. Such behaviour was attributed to stabilization of coil1A dimer due to mutation. Indeed, this residue is in a 'd' position where the small hydrophobic cysteine side chain is preferred over tyrosine. These authors observed partial formation of a disulphide bond by the mutated residue, although the negative effect of the mutation on Acn interaction was also observed in reducing conditions. Parallel observations were made in the past for vimentin coil1A region which was demonstrated to be only marginally stable [30]. Moreover, point mutation Y117L in vimentin coil1A (i.e., in a position equivalent to Y45 in LA) resulted in a loss of longitudinal assembly, indicating a key role of coil1A unzipping in the latter process [30]. More recently another vimentin mutation Y400L near the C-terminus of the rod was shown to have the same effect, suggesting that also the end of coil2 should be unzipping towards longitudinal assembly [74].

As the next step, using SEC we could show that the pairs of N- and C-terminal LA fragments stabilized by fusions with capping motifs indeed interact in solution. Our observations are in line with earlier studies using longer lamin fragments without capping motifs [40]. In particular, we could confirm that the Acn interaction appears to be relatively weak in solution. Indeed, our SEC-MALS data hinted towards a dynamic equilibrium between individual fragments and the complex (Figure 6).

Finally, we applied chemical cross-linking to explore the 3D architecture of the heterotetrameric Acn complex. In line with the demonstrated role of N- and C-terminal unzipping of the rod domain in assembly, our starting hypothesis towards the three-dimensional modelling of the complex was a parallel CC heterotetramer (dimer of dimers), following a suggestion outlined in [64]. We adjusted the overlap of the two dimers to maximize the fit to the chemical cross-links obtained, which resulted in an overlap of 6.5 nm (Figure 8). Of note, different values for the Acn overlap have been suggested in the past. Early electron microscopy data for in vitro assembled lamin threads visualized using rotary metal shadowing suggested a 1-3 nm overlap [24], while recent cryo-EM studies of natively assembled lamin filaments are more compatible with an overlap of 10 nm [23]. In each case the overlap was estimated based on the expected length of lamin rod (50-51 nm) and the observed periodic appearance of the globular tail domains along the filament. A recent study [15] suggested an Acn overlap of 14 nm, as calculated on the basis of a chemical cross-link between LA 1-300 and LA 286-400 (R388C) fragments and the A11 overlap known from crystallographic studies [12,15]. Most recently, chemical cross-linking studies of full-length human LA have yielded the Acn overlap value of 5-6nm [70], which is in accordance with our result.

To our knowledge, our heterotetrameric model is the first attempt to approach the Acn interaction in molecular detail. Our data highlight the specific regions that are critical for the correct longitudinal assembly of lamin filaments and indicate that this assembly depends on a delicate

balance of molecular interactions. These observations explain the fact that lamin mutations (such as the ones discussed above) located in the N- and C-terminal regions of the rod can be detrimental to the assembly, ultimately leading to the disease phenotype. At the same time, the signature globular domain (residues 434–552) present within the lamin A tail is known to be the hotspot of interactions between lamina and its protein partners such as actin [75] and BAF1 (barrier-to-autointegration factor) which further mediates the interaction with LAP2α [5] and emerin [76]. Since the globular tail domain is connected to the rod by a flexible region (as confirmed by the recent cryoEM data [23]) one should expect that the interactions with partners are preserved also when individual lamin dimers are incorporated into the lamina. Altogether, over 50 lamin-associated proteins have been described but for a majority them the details of the interaction are still elusive [77].

In the future, further efforts towards a structure with true atomic precision, using either crystallography or cryo-EM, should be undertaken. Indeed, it is hoped that additional Acn complexes, based on newly designed N- and C-terminal fragments and/or additionally stabilized complexes (e.g., through cross-linking) could be prone to crystallization. At the same time, cryo-EM studies could employ longer constructs and/or full-length proteins.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4409/9/7/1633/s1, Supplementary Table 1: DNA primers used, Supplementary Table 2: X-ray diffraction and refinement statistics, Supplementary Table 3: Chemical cross-links obtained for the complexes of N- and C-terminal fragments, Supplementary Table 4: Properties of the final Acn model as analysed using program PISA, Supplementary Figure 1: Sequence overview and comparison with related lamins and IF proteins, Supplementary Figure 2: Flowchart of molecular modelling and refinement of the heterotetramer, Supplementary Figure 3: Analysis of the Gp7$_{F40C}$-LA 327-403 structure, Supplementary Figure 4: Crystal structure of the LA 17-70-Eb1 fusion, Supplementary Figure 5: Topological scheme of the heterotetramer involving an overlap of partially unzipped N- and C-terminal dimers. Supplementary Figure 6: Possible higher-order association of two Acn tetramers, Supplementary Text 1: Detailed information for LC-MS/MS analysis.

Author Contributions: Conceptualization, G.S., A.V.L. and S.V.S.; investigation, G.S., A.V.L., P-J.V. and J.F.; writing—original draft preparation, G.S. and A.V.L.; writing—review and editing, G.S., A.V.L., P-J.V., P.N. and S.V.S.; funding acquisition, S.V.S. and P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research (including the APC) was funded by the CELSA alliance grant number 18/044 and European Commission H2020 (project EU_FT-ICR_MS - grant agreement ID: 731077).

Acknowledgments: We thank the teams of beamline I04 at Diamond Light Source (UK), beamline Proxima-1 at Synchrotron Soleil (France) and beamline P14 at Petra III storage ring (Germany) as well as Dr. Evgenii Osipov for help with crystallographic data collection. We are also grateful to the members VIB Switch Laboratory for help with the SEC-MALS instrument which is part of the Molecular Biophysics platform of the KU Leuven. Lastly, we acknowledge the Centre of Molecular Structure Core Facility at BIOCEV, a Structural Mass Spectrometry Facility funded by European Regional Development Funds (CZ.1.05/1.1.00/02.0109 BIOCEV) and supported by the Czech Infrastructure for Integrative Structural Biology (LM2018127 CIISB for CMS BIOCEV funded by MEYS CR).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Turgay, Y.; Medalia, O. The structure of lamin filaments in somatic cells as revealed by cryo-electron tomography. *Nucleus* **2017**, *8*, 475–481.
2. Collas, P.; Lund, E.G.; Oldenburg, A.R. Closing the (nuclear) envelope on the genome: How nuclear lamins interact with promoters and modulate gene expression. *BioEssays* **2014**, *36*, 75–83.
3. Frock, R.L.; Kudlow, B.A.; Evans, A.M.; Jameson, S.A.; Hauschka, S.D.; Kennedy, B.K. Lamin A/C and emerin are critical for skeletal muscle satellite cell differentiation. *Genes Dev.* **2006**, *20*, 486–500.
4. Maynard, S.; Keijzers, G.; Akbari, M.; Ezra, M. Ben; Hall, A.; Morevati, M.; Scheibye-Knudsen, M.; Gonzalo, S.; Bartek, J.; Bohr, V.A. Lamin A/C promotes DNA base excision repair. *Nucleic Acids Res.* **2019**,

47, 11709–11728.

5. Qi, R.; Xu, N.; Wang, G.; Ren, H.; Li, S.; Lei, J.; Lin, Q.; Wang, L.; Gu, X.; Zhang, H.; et al. The lamin-A/C-LAP2α-BAF1 protein complex regulates mitotic spindle assembly and positioning. *J. Cell Sci.* **2015**, *128*, 2830–2841.
6. Shumaker, D.K.; Solimando, L.; Sengupta, K.; Shimi, T.; Adam, S.A.; Grunwald, A.; Strelkov, S. V.; Aebi, U.; Cardoso, M.C.; Goldman, R.D. The highly conserved nuclear lamin Ig-fold binds to PCNA: Its role in DNA replication. *J. Cell Biol.* **2008**, *181*, 269–280.
7. Shimi, T.; Pfleghaar, K.; Kojima, S.I.; Pack, C.G.; Solovei, I.; Goldman, A.E.; Adam, S.A.; Shumaker, D.K.; Kinjo, M.; Cremer, T.; et al. The A- and B-type nuclear lamin networks: Microdomains involved in chromatin organization and transcription. *Genes Dev.* **2008**, *22*, 3409–3421.
8. Kang, S. mi; Yoon, M.H.; Park, B.J. Laminopathies; Mutations on single gene and various human genetic diseases. *BMB Rep.* 2018, *51*, 327–337.
9. Broers, J.L.V.; Ramaekers, F.C.S. The role of the nuclear lamina in cancer and apoptosis. *Adv. Exp. Med. Biol.* **2014**, *773*, 27–48.
10. de Toledo, M.; Lopez-Mejia, I.C.; Cavelier, P.; Pratlong, M.; Barrachina, C.; Gromada, X.; Annicotte, J.-S.; Tazi, J.; Chavey, C. Lamin C Counteracts Glucose Intolerance in Aging, Obesity and Diabetes Through β-Cell Adaptation. *Diabetes*, **2020**, *69*, 647–660.
11. Pradhan, R.; Jayakrishnan Nallappa, M.; Sengupta, K. Lamin A/C modulates spatial organization and function of the Hsp70 gene locus via Nuclear Myosin I (NM1). *J.Cell. Sci.*, **2020**, *133*, jcs:236265.
12. Lilina, A. V.; Chernyatina, A.A.; Guzenko, D.; Strelkov, S. V. Lateral A11 type tetramerization in lamins. *J. Struct. Biol.* **2020**, *209*, 107404.
13. Lupas, A.N.; Bassler, J. Coiled Coils – A Model System for the 21st Century. *Trends Biochem. Sci.* 2017, *42*, 130–140.
14. Chernyatina, A.A.; Guzenko, D.; Strelkov, S. V. Intermediate filament structure: The bottom-up approach. *Curr. Opin. Cell Biol.* 2015, *32*, 65–72.
15. Ahn, J.; Jo, I.; Kang, S. mi; Hong, S.; Kim, S.; Jeong, S.; Kim, Y.H.; Park, B.J.; Ha, N.C. Structural basis for lamin assembly at the molecular level. *Nat. Commun.* **2019**, *10*, 3757.
16. WEBER, K.; GEISLER, N. Intermediate Filaments: Structural Conservation and Divergence. *Ann. N. Y. Acad. Sci.* **1985**, *455*, 126–143.
17. Heitlinger, E.; Peter, M.; Lustig, A.; Villiger, W.; Nigg, E.A.; Aebi, U. The role of the head and tail domain in lamin structure and assembly: Analysis of bacterially expressed chicken Lamin A and truncated B2 lamins. *J. Struct. Biol.* **1992**, *108*, 74–91.
18. Hess, J.F.; Budamagunta, M.S.; Aziz, A.; FitzGerald, P.G.; Voss, J.C. Electron paramagnetic resonance analysis of the vimentin tail domain reveals points of order in a largely disordered region and conformational adaptation upon filament assembly. *Protein Sci.* **2013**, *22*, 47–55.
19. Lee, C.H.; Coulombe, P.A. Self-organization of keratin intermediate filaments into cross-linked networks. *J. Cell Biol.* **2009**, *186*, 409–421.
20. Geisler, N.; Schünemann, J.; Weber, K.; Häner, M.; Aebi, U. Assembly and architecture of invertebrate cytoplasmic intermediate filaments reconcile features of vertebrate cytoplasmic and nuclear lamin-type intermediate filaments. *J. Mol. Biol.* **1998**, *282*, 601–617.
21. Heitlinger, E.; Peter, M.; Lustig, A.; Nigg, E.A. Expression of Chicken Lamin B2 in Escherichia coli: Characterization of its Structure, Assembly, and Molecular Interactions. *Cell* **1991**, *113*, 485–495.

22. Aebi, U.; Julie, C.; Buble, L.; Gerace, L. The nuclear lamina is a meshwork of intermediate-type filaments. *Nature* **1986**, *324*, 698–699.
23. Turgay, Y.; Eibauer, M.; Goldman, A.E.; Shimi, T.; Khayat, M.; Ben-Harush, K.; Dubrovsky-Gaupp, A.; Sapra, K.T.; Goldman, R.D.; Medalia, O. The molecular architecture of lamins in somatic cells. *Nature* **2017**, *543*, 261–264.
24. Stuurman, N.; Heins, S.; Aebi, U.; Mü, M.E. Nuclear Lamins: Their Structure, Assembly, and Interactions. *J. Struct. Biol.* **1998**, *122*, 42–66.
25. Ben-Harush, K.; Wiesel, N.; Frenkiel-Krispin, D.; Moeller, D.; Soreq, E.; Aebi, U.; Herrmann, H.; Gruenbaum, Y.; Medalia, O. The Supramolecular Organization of the C. elegans Nuclear Lamin Filament. *J. Mol. Biol.* **2009**, *386*, 1392–1402.
26. Foeger, N.; Wiesel, N.; Lotsch, D.; Mücke, N.; Kreplak, L.; Aebi, U.; Gruenbaum, Y.; Herrmann, H. Solubility properties and specific assembly pathways of the B-type lamin from Caenorhabditis elegans. *J. Struct. Biol.* **2006**, *155*, 340–350.
27. ISOBE, K.; GOHARA, R.; UEDA, T.; TAKASAKI, Y.; ANDO, S. The Last Twenty Residues in the Head Domain of Mouse Lamin A Contain Important Structural Elements for Formation of Head-to-Tail Polymers *in Vitro*. *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1252–1259.
28. Strelkov, S. V.; Herrmann, H.; Geisler, N.; Lustig, A.; Ivaninskii, S.; Zimbelmann, R.; Burkhard, P.; Aebi, U. Divide-and-conquer crystallographic approach towards an atomic structure of intermediate filaments. *J. Mol. Biol.* **2001**, *306*, 773–781.
29. Chernyatina, A.A.; Hess, J.F.; Guzenko, D.; Voss, J.C.; Strelkov, S. V. How to Study Intermediate Filaments in Atomic Detail. *Methods Enzymol.* **2016**, *568*, 3–33.
30. Meier, M.; Padilla, G.P.; Herrmann, H.; Wedig, T.; Hergt, M.; Patel, T.R.; Stetefeld, J.; Aebi, U.; Burkhard, P. Vimentin Coil 1A-A Molecular Switch Involved in the Initiation of Filament Elongation. *J. Mol. Biol.* **2009**, *390*, 245–261.
31. Strelkov, S. V.; Herrmann, H.; Geisler, N.; Wedig, T.; Zimbelmann, R.; Aebi, U.; Burkhard, P. Conserved segments 1A and 2B of the intermediate filament dimer: Their atomic structures and role in filament assembly. *EMBO J.* **2002**, *21*, 1255–1266.
32. Kapinos, L.E.; Burkhard, P.; Herrmann, H.; Aebi, U.; Strelkov, S. V; Müller, M. Simultaneous Formation of Right- and Left-handed Anti-parallel Coiled-coil Interfaces by a Coil2 Fragment of Human Lamin A. *J. Mol. Biol* **2011**, *408*, 135–146.
33. Tao, Y.; Strelkov, S. V.; Mesyanzhinov, V. V.; Rossmann, M.G. Structure of bacteriophage T4 fibritin: A segmented coiled coil and the role of the C-terminal domain. *Structure* **1997**, *5*, 789–798.
34. O'Shea, E.K.; Klemm, J.D.; Kim, P.S.; Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science (80-.).* **1991**, *254*, 539–544.
35. Morais, M.C.; Kanamarul, S.; Badasso, M.O.; Koti, J.S.; Owen, B.A.L.; McMurray, C.T.; Anderson, D.L.; Rossmann, M.G. Bacteriophage φ29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Biol.* **2003**, *10*, 572–576.
36. Slep, K.C.; Rogers, S.L.; Elliott, S.L.; Ohkura, H.; Kolodziej, P.A.; Vale, R.D. Structural determinants for EB1-mediated recruitment of APC and spectraplakins to the microtubule plus end. *J. Cell Biol.* **2005**, *168*, 587–598.
37. Frye, J.; Klenchin, V.A.; Rayment, I. Structure of the tropomyosin overlap complex from chicken smooth muscle: Insight into the diversity of N-terminal recognition. *Biochemistry* **2010**, *49*, 4908–4920.

38. Taylor, K.C.; Buvoli, M.; Korkmaz, E.N.; Buvoli, A.; Zheng, Y.; Heinze, N.T.; Cui, Q.; Leinwand, L. a.; Rayment, I. Skip residues modulate the structural properties of the myosin rod and guide thick filament assembly. *Proc. Natl. Acad. Sci.* **2015**, *112*, E3806–E3815.
39. Korkmaz, E.N.; Taylor, K.C.; Andreas, M.P.; Ajay, G.; Heinze, N.T.; Cui, Q.; Rayment, I. A composite approach towards a complete model of the myosin rod. *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 172–189.
40. Kapinos, L.E.; Schumacher, J.; Mücke, N.; Machaidze, G.; Burkhard, P.; Aebi, U.; Strelkov, S. V.; Herrmann, H. Characterization of the Head-to-Tail Overlap Complexes Formed by Human Lamin A, B1 and B2 'Half-minilamin' Dimers. *J. Mol. Biol.* **2010**, *396*, 719–731.
41. Kochin, V.; Shimi, T.; Torvaldson, E.; Adam, S.A.; Goldman, A.; Pack, C.G.; Melo-Cardenas, J.; Imanishi, S.Y.; Goldman, R.D.; Eriksson, J.E. Interphase phosphorylation of lamin A. *J. Cell Sci.* **2014**, *127*, 2683–2696.
42. Weeks, S.D.; Drinker, M.; Loll, P.J. Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr. Purif.* **2007**, *53*, 40–50.
43. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234.
44. Studier, F.W. Stable expression clones and auto-induction for protein production in E. Coli. *Methods Mol. Biol.* **2014**, *1091*, 17–32.
45. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132.
46. Panjikar, S.; Parthasarathy, V.; Lamzin, V.S.; Weiss, M.S.; Tucker, P.A. Auto-Rickshaw: An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61*, 449–457.
47. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.* 2008, *64*, 112–122.
48. Read, R.J.; McCoy, A.J. Using SAD data in Phaser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 338-344.
49. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 1002–1011.
50. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186–190.
51. Winter, G.; Waterman, D.G.; Parkhurst, J.M.; Brewster, A.S.; Gildea, R.J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I.D.; et al. DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.* **2018**, *74*, 85–97.
52. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 72–82.
53. Evans, P.R.; Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1204–1214.
54. Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 22–25.
55. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.
56. Vagin, A.A.; Steiner, R.A.; Lebedev, A.A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G.N. REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta*

Crystallogr. Sect. D Biol. Crystallogr. **2004**, *60*, 2184–2195.

57. Bricogne, G.; Blanc, E.; Brand, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O.S.; Vonrhein, C.; et al. BUSTER. *Cambridge, United Kingdom Glob. Phasing Ltd.* 2017.
58. Strelkov, S. V.; Burkhard, P. Analysis of α-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **2002**, *137*, 54–64.
59. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G.M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P.G. Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **2004**, *25*, 1327–1333.
60. Rozbeský, D.; Rosůlek, M.; Kukačka, Z.; Chmelík, J.; Man, P.; Novák, P. Impact of Chemical Cross-Linking on Protein Structure and Function. *Anal. Chem.* **2018**, *90*, 1104–1113.
61. Fiala, J.; Kukačka, Z.; Novák, P. Influence of cross-linker polarity on selectivity towards lysine side chains. *J. Proteomics* **2020**, *218*.
62. Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C.H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. StavroX-A software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 76–87.
63. Iacobucci, C.; Götze, M.; Ihling, C.H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. *Nat. Protoc.* **2018**, *13*, 2864–2889.
64. Strelkov, S. V; Schumacher, J.; Burkhard, P.; Aebi, U.; Herrmann, H.; Mu, M.E. Crystal Structure of the Human Lamin A Coil 2B Dimer : Implications for the Head-to-tail Association of Nuclear Lamins. *J. Mol. Biol.* **2004**, *343*, 1067–1080.
65. Heo, L.; Lee, H.; Seok, C. GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. *Sci. Rep.* **2016**, *6*, 1–10.
66. Ko, J.; Park, H.; Heo, L.; Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **2012**, *40*, 294–297.
67. Kraatz, S.H.W.; Bianchi, S.; Steinmetz, M.O. Combinatorial use of disulfide bridges and native sulfur-SAD phasing for rapid structure determination of coiled-coils. *Biosci. Rep.* **2018**, *38*, 1–11.
68. Guzenko, D.; Strelkov, S. V. Optimal data-driven parameterization of coiled coils. *J. Struct. Biol.* **2018**, *204*, 125–129.
69. Doig, A.J.; Baldwin, R.L. N- and C-capping preferences for all 20 amino acids in α-helical peptides. *Protein Sci.* **1995**, *4*, 1325–1336.
70. Makarov, A.A.; Zou, J.; Houston, D.R.; Spanos, C.; Solovyova, A.S.; Cardenal-Peralta, C.; Rappsilber, J.; Schirmer, E.C. Lamin A molecular compression and sliding as mechanisms behind nucleoskeleton elasticity. *Nat. Commun.* **2019**, *10*, 3056.
71. Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **2011**, *173*, 530–540.
72. Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. **2007**, *774–797*.
73. Herrmann, H.; Aebi, U. Intermediate Filaments: Molecular Structure, Assembly Mechanism, and Integration Into Functionally Distinct Intracellular Scaffolds. *Annu. Rev. Biochem.* **2004**, *73*, 749–789.
74. Kaus-Drobek, M.; Mücke, N.; Szczepanowski, R.H.; Wedig, T.; Czarnocki-Cieciura, M.; Polakowska, M.; Herrmann, H.; Wysłouch-Cieszyńska, A.; Dadlez, M. Vimentin S-glutathionylation at Cys328 inhibits

filament elongation and induces severing of mature filaments in vitro. *FEBS J.* **2020**, doi:10.1111/febs.15321.

75. Simon, D.N.; Zastrow, M.S.; Wilson, K.L. Direct actin binding to A- and B-type lamin tails and actin filament bundling by the lamin A tail. *Nucleus* **2010**, *1*, 264–272.
76. Samson, C.; Petitalot, A.; Celli, F.; Herrada, I.; Ropars, V.; Le Du, M.-H.; Nhiri, N.; Jacquet, E.; Arteni, A.-A.; Buendia, B.; et al. Structural analysis of the ternary complex between lamin A/C, BAF and emerin identifies an interface disrupted in autosomal recessive progeroid diseases. *Nucleic Acids Res.* **2018**, *46*, 10460–10473.
77. Simon, D.N.; Wilson, K.L. Partners and post-translational modifications of nuclear lamins. *Chromosoma* **2013**, *122*, 13–31.